Isolation and characterisation of uncommon myxobacteria and other marine bacteria and their multilocus sequences analyses from Indonesia's biodiversity

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades

einer Doktorin der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von Rina Andriyani
aus Bandung / Indonesia

1. Referentin oder Referent: Professor Dr. Michael Steinert
2. Referentin oder Referent: Privatdozent Dr. Joachim Wink
eingereicht am: 01.03.2021
mündliche Prüfung (Disputation) am: 20.05.2021

Druckjahr 2021

The Entirely Merciful, the Especially Merciful

This dissertation is dedicated to my beloved parents and our big family

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Posterbeiträge

1. R. Andriyani, K. I. Mohr, T. Mozef, M. Steinert, J. Wink, "Screening of Indonesian myxobacteria for new antibacterials and antifungals", The 7th HIPS Symposium, Saarbrücken, 2017.

2. R. Andriyani, E. Sudarman, K. I. Mohr, T. Mozef, M. Steinert, J. Wink, "Compounds from *Myxococcus* sp. (MxGP53TM) a strain isolated from Indonesian soil", The 10th International PhD Symposium, HZI, Braunschweig, 2017.

3. R. Andriyani, K. I. Mohr, T. Mozef, M. Steinert, J. Wink, et al., "A novel gliding bacterium of the family *Flammeovirgaceae* isolated from Bali Beach Indonesia", The 11th International PhD Symposium, HZI Braunschweig, 2018.

4. R. Andriyani, K. I. Mohr, C. Risdian, T. Mozef, M. Steinert, J. Wink, "Drug discovery pipeline from a gliding bacterium to the isolation of antibacterial compounds by utilizing Indonesian biodiversity", The 12th International PhD Symposium, HZI Braunschweig, 2019.

Preface and acknowledgements

First of all, I would like to thank all of German-Indonesian Antiinfective Cooperation's stakeholders in Germany and Indonesia. Thanks to the Federal Ministry of Education and Research, the Helmholtz Centre for Infection Research, the Graduate School of Helmholtz Centre for Infection Research and the Technische Universität Braunschweig in Germany. Thanks to the Ministry of Research and Technology of Indonesia, the Research Center of Chemistry and the Research Unit for Cleaning Technology in Indonesia.

To Dr. Tjandrawati Mozef as the German-Indonesian Antiinfective Cooperation project's lead collaborator from Indonesia who provided this opportunity and supported me. To the Head of Indonesian Institute of Sciences (LIPI), Dr. L.T. Handoko, the Head of Research Center of Chemistry at that time, Dr. Agus Haryono, the Head of Research Unit for Cleaning Technology (previous and present) Dr. Rike Yudianti and Dr. Ajeng Arum Sari, who gave me the eligibility for receiving this duty. Thanks to the staff as well.

To the Director of Helmholtz Centre for Infection Research, the Director of the Graduate School of Helmholtz Centre for Infection Research, the Dean of The Faculty of Life Sciences of Technische Universität Braunschweig, the Head of Microbiology Department of Technische Universität Braunschweig and the staff as well who have assisted me for administration business during this study.

Special thanks to my supervisors and thesis committee members, PD Dr. Joachim Wink and Prof. Dr. Michael Steinert. Without the Letter of Acceptance from you, it would have been impossible to receive the Riset-Pro scholarship from the Indonesian government to enrol in the PhD programme in the Technische Universität Braunschweig and the Graduate School of Helmholtz Centre for Infection Research. Also, to Dr. Kathrin I. Mohr and Dr. Raimo Franke as two of the thesis committee members. Thank you for your advice, support, keywords and guidance for my research.

To Prof. Marc Stadler as the Department of Microbial Drugs (MWIS) group leader and the staff for allowing me to work in the chemistry laboratory. To the team of Microbial Strain Collection (MISG): Stephanie Schulz, Klaus P. Conrad, Wera Collisi, Jolla Lulla, Birte Trunkwalter and Aileen Gollasch. To all the PhD students, the internship students and the postdoctoral researchers in the Helmholtz Centre for Infection Research, especially in the MISG and MWIS groups, also the PhD students from the Indonesian Institute of Sciences during this period. Thank you for the discussion moments and the togetherness on special occasions.

To all teachers and lecturers who have shared their knowledge and broadened my perspective. It is not only about the science; thank you very much for your dedication and professionalism.

To all friends everywhere, especially Christina Pape, Rani, Fitri, Okky, Yani, Rossy Chaerun Nisa, Rahma, Amelia, Mba Rere, Ulrike Tarajona, Mango, Uthay, Florian, Perda, Kikin, Astri, Khadija, Kwame, Tanya, Indonesian Moeslem Union of Braunschweig, and Indonesian Students Union of Braunschweig. Thank you for your support and friendship.

Thanks to all authors in my references and Google, which made everything easier to find. I would like to thank Enago (www.enago.com) for the English language review.

Last but not least, thanks to the Greatest who makes everything possible, has been guiding my soul to think clearly when I felt lost and doubtful. You always send answers, even in an unpredictable way. I have known You deeply in every chapter of my life.

CONTENTS

Abbreviations

°C	degree of Celsius
°E	degree of East
°N	degree of North
°S	degree of South
µL	mikroliter
A	adenine
Art-Nr.	article number
BLAST	Basic Local Alignment Search Tool
BMBF	*Bundesministerium für Bildung und Forschung*
C	cytosine
CDS	Coding sequence
cm	centimetre
DAAD	*Deutscher Akademischer Austauschdienst*
DNA	deoxyribonucleic acid
Dr.	Doctor
Dr.rer.nat	*Doctor rerum naturalium*
DSMZ	*Deutsche Sammlung von Mikroorganismen und zellkulturen*
e.g.	*exempli gratia*
et al.	*et alia*
etc.	*et cetera*
F	forward

FA(s)	fatty acid(s)
G	guanine
G + C	guanine + cytosine
GC-MS	Gas Chromatography-Mass Spectrometry
GMAK	*Genomeanalytik*
HIPS	Helmholtz Institute for Pharmaceutical Research Saarland
HPLC	High Performance Liquid Chromatography
HRESIMS	High-Resolution Electron Spray Ionisation Mass Spectrometry
HZI	*Helmholtz Zentrum für Infektionsforschung*
i.e.	*id est*
LIPI	*Lembaga Ilmu Pengetahuan Indonesia*
MEGA	Molecular Evolutionary Genetics Analysis
min	minute
MISG	*Mikrobielle Stammsammlung Gruppe*
mL	millilitre
mm	millimetres
Mt.	Mountain
MWIS	*Mikrobielle Wirkstoffe*
NCBI	National Center for Biotechnology Information
ng	nanogram
nm	nanometer
PD Dr.	*Privatdozent Doctor*

pH	potential of hydrogen
PhD	*Philosophiae Doctor*
pmol	picomole, equal with 10^{-12} moles
Prof. Dr.	Professor Doctor
R	reverse
RAST	Rapid Annotations using Subsystems Technology
Riset-pro	Research and Innovation In Science and Technology Project
RNA	ribonucleic acid
RP-HPLC	Reverse-Phase High Performance Liquid Chromatography
rpm	rotary per minute
rRNA	ribosomal Ribonucleic acid
s	second
S	svedberg unit, is a measure of sedimentation velocity and mass
Spp.	species plural
T	thymine
t-RNA	transfer-Ribonucleic acid
TU	*Technische Universität*

Chapter 1. Introduction

1.1 Background

Antimicrobial resistance is a growing problem around the world. Resistance develops when antimicrobials' daily dose becomes ineffective against the microbial target, leaving several microbes with resistant genes. These resistant microbes can then multiply and spread to a global scale. The main threat from resistant microbes is that their presence increases the risk of fatal outcomes in infected populations.

The World Health Organisation has begun to take this issue of antimicrobial resistance seriously. Its *Global Antimicrobial Resistance and Use Surveillance System Report: Early Implementation 2020* reported resistant microbes in blood and urine samples from all participating countries. The investigation that formed the basis for this report was conducted in 2018 and revealed that Indonesia (as one of the participating countries) exhibited infections of *Acinetobacter* spp., *Escherichia coli*, *Klebsiella pneumonia*, *Staphylococcus aureus*, *Streptococcus pneumonia*, and *Salmonella*. These cases also showed that some antibiotics had become ineffective against these bacteria in Indonesia. The report further highlighted that *Escherichia coli* is the most common resistant bacterium in both

Reports about antimicrobial resistance is an urgent need for research into novel antimicrobial agents. With its wealth of biodiversity, Indonesia has the resources to find a solution to this global problem. At the same time, research on antimicrobials resistance requires international collaboration among institutions all around the world. Indonesia and Germany share a common interest on antimicrobial resistance. From 2015 to 2019, the two countries conducted joint research under GINAICO, involving LIPI and the HZI with additional support from the HZI Graduate School and the TU Braunschweig doctoral programme.

1.2 Aims

This study aims to isolate novel and/or rare antimicrobial substances from gliding bacteria (especially myxobacteria), identify active compounds that exhibit antimicrobial activity and study the taxonomy of gliding bacteria.

1.3 Methodology

The methods to achieve this research aim to isolate and identify gliding bacteria, screen for antiinfectives against Gram-positive bacteria, Gram-negative bacteria, and fungi, identify active compounds and conduct a polyphasic taxonomy analysis for a new strain using classical and modern taxonomy.

1.4 Research location

This study forms part of the cooperative effort to utilise samples from several regions in Indonesia. The research was conducted at the research group's microbial strain collection (MISG) at the HZI from 2016 to 2020.

1.5 Research funding

Funding was provided by the DAAD and the BMBF. Additionally, scholarships were sponsored by Riset-Pro, Indonesia from October 2016 to October 2020 and the HZI Graduate School from November 2020 to December 2020.

Chapter 2. Screening of culturable gliding bacteria isolated from some regions in Indonesia as sources for antiinfective substances

Indonesia is a country with immense potential for biodiversity exploration, especially in the field of microorganisms. In this study, a rare Myxobacterium and a novel non-myxobacterial organism were isolated. We also screened for antiinfectives against Escherichia coli, Candida albicans and Staphylococcus aureus. Also, we conducted a secondary metabolite analysis using the High-Pressure Liquid Chromatography and High-Resolution Electrospray Ionisation Mass Spectrometry (HRESIMS) methods.

2.1 Background

Indonesia is an archipelagic state with some 18,808 islands [1], the largest of which are Sumatra, Java, Borneo, Sulawesi, Bali, and Papua. The country stretches between longitudes 97°E and 141°E, and latitudes 6°N and 11°S [1]. Indonesia consists of 2.8 million square kilometres of water (covering 92,877 square kilometres of inland waters) and 1,826,440 square kilometres of land [1]. The geomorphology of Indonesia is responsible for regional differences in climate within the country. Climate is defined as the average weather pattern over 30 years or more and measured in parameters such as temperature, air pressure, wind, rainfall and humidity [2]. The samples for this research came from several Indonesian regions, such as Jakarta, West Java, East Java, South-eastern Sulawesi, and Bali. Every sampling area in this study has a characteristic climate. Generally speaking, the average rainfall in Java and Bali is 2500 mm per year, while in Sulawesi, it varies widely between 54 and 6003 mm per year [2]. Furthermore, the average temperature in Java is between 21–34°C [2], while Sulawesi has a distinct climate compared to the other regions with an average temperature of 24.6–28.7°C and average air pressure of 1004.5–1013.3 millibars [2]. The average humidity in Sulawesi lies within the range of 65%–90% [2]. We assumed that these differences in climate would influence the variety of the bacteria we isolated.

Bacteria constitute the domain prokaryote and need appropriate habitats to grow. Soil, seawater and sedimentary soil subsurface are naturally the three largest

habitats where bacteria could grow [3]. Soil is one of the best bacterial growth media because it is a major reservoir of macronutrients such as carbon, nitrogen and 25 other chemical elements [3–7]. Each prokaryote (bacterial) species has a different optimal growth condition, making their growth dependent upon environmental factors such as pH [6,8].

Prokaryotes are cells with co-transcriptional translation on their main chromosomes. They translate nascent messenger RNAs into protein and possess transcriptionally coupled translation [9]. Since the early1970's, Carl Woese has been using a prokaryote classification method based on the RNA component of small subunit ribosomes (SSU), henceforth known as SSU rRNA or 16S rRNA [10]. This method uses SSU rRNA gene sequences' similarities to discern evolutionary relationships among prokaryotes (Bacteria and Archaea) [10].

One of the characteristics used to select the bacteria targeted in this research is that they should not be pathogens. Pathogenicity is the ability of bacteria to cause disease in a host [8]. Bacteria are classified into pathogens and non-pathogens based on their activity patterns. Pathogens establish an infection when they (a) colonise a host; (b) replicate in host tissues once they find a nutritionally compatible niche; (c) defeats the host's defences and adaptive immune response; (d) impair the host and (e) exit and spread to a new host [8,11].

Aside from their host activities, bacteria can be divided based on cell wall staining into Gram-positive and Gram-negative types. Kapoor (2017) [12] wrote about the differences in bacterial cells' basic anatomy between Gram-positive and Gram-negative bacteria. The cell walls of Gram-positive bacteria consist of a tough and rigid mesh–the cytoplasmic membrane. Meanwhile, Gram-negative bacteria possess a thin cell wall with a separate outer membrane (second lipid membrane). Gram-negative bacteria have a periplasm separating the outer membrane from the cytoplasmic membrane. The outer membrane in Gram-negative bacteria is a protective layer that defends the bacterium from many substances. This membrane contains porins, which are channels that allow the controlled ingress of various molecules, including drugs. The cell wall gives the bacterium a characteristic shape

and protects it from osmotic and mechanical stresses. The cytoplasmic membrane restricts ionic fluxes into or out of the cell and keeps the cytoplasmic and bacterial components in a settled space [12]. Gram-positive and Gram-negative bacteria and fungi are targets for antibacterial assays to look for candidate substances. Some of the microbes used in this role are *Staphylococcus aureus*, *Escherichia coli* and *Candida albicans*. All of these pathogens have previously been found with resistance genes.

Pathogens contain concrete virulence genes that act as intermediaries in interactions with the host, deriving particular responses from the host cell that facilitate the replication and dissemination of the pathogen [11]. Both pathogenic and non-pathogenic bacteria produce secondary metabolites as antimicrobials to protect themselves from hostile substances and organisms that try to pass their defences. Antimicrobials are organic compounds (whether natural or synthetic) that inhibit or destroy particular bacteria, preferably even at low concentrations [13]. Antimicrobials' possible action mechanisms are targeting the cell membrane or cell wall, inhibiting protein biosynthesis, inhibiting DNA replication and inhibiting folic acid metabolism [12,14].

Antimicrobials targeting the cell membrane. The cell envelope (cell membrane and cell wall) protects the organism from hostile environmental factors such as extreme osmolarity, harsh chemicals and even antimicrobials [13]. The cell membrane or cytoplasmic membrane of bacteria is composed of phospholipids and proteins (70% of the membrane's mass) that can be synthesised in the cell or taken from the environment as building blocks [13,14]. The main targets for antimicrobials in cell membranes are the metabolic steps of FA synthesis and membrane phospholipids [14]. The major functions of the cytoplasmic membrane are (i) selective permeability and transport of solutes; (ii) electron transport and oxidative phosphorylation in aerobic species; (iii) excretion of hydrolytic exoenzymes; (iv) bearing the enzymes and vehicle molecules that function in the biosynthesis of DNA, cell wall polymers, and membrane lipids and (v) centre of the receptors and the chemotactic proteins and sensory transduction systems [13].

Antimicrobials targeting the cell wall. A microorganism's cell walls are built of substances such as peptidoglycan, murein, or mucopeptide (all are synonyms) [13,14]. As protection against osmotic stresses, the cell wall plays a role in cell division that is essential to biosynthesis [13]. Gram-negative bacteria's cell wall surface contains sites that function as major antigenic determinants and one lipopolysaccharide component is responsible for the nonspecific endotoxin activity of Gram-negative bacteria [13]. The cell wall acts as a corseting structure (best characterised as a fishing net), protecting the cell against osmotic lysis. Therefore, substances that impair the wall (e.g. lysozyme, which cleaves the sugar linkages) or inhibit its normal synthesis (e.g. penicillin, which interrupts peptidyl cross-linkages) can cause cell lysis [13].

Antimicrobials inhibiting the biosynthesis of protein. Protein biosynthesis takes place in the ribosome as the main cellular apparatus for this process [11,15]. A bacterial ribosome consists of 70S particles that can be divided into a 30S SSU (Small subunit) and a 50S LSU (Large subunit) [13,15,16]. The 50S subunit contains a 23S rRNA (2,900 nucleotides) and a 5S rRNA (120 nucleotides) in a complex with 33 proteins, whereas the 30S subunit contains 16S rRNA (1540 nucleotides) bonding with 21 proteins [15,17–20]. Outer membrane proteins are synthesised on ribosomes bound to the cell membrane's cytoplasmic surface [13]. How they are transferred to the outer membrane is still uncertain, but one hypothesis suggests that transfer occurs at zones of adhesion between the cytoplasmic and outer membranes, visible in an electron microscope [13]. Protein synthesis starts with the ribosome subunits coming together, remaining together as individual amino acids are added to the end of the growing peptide, and then separating again in conjunction with the finished protein release [17]. The ribosome’s subunits, its chemical composition and its functional specificities are sufficiently distinct to explain why antimicrobial drugs can inhibit protein synthesis in bacterial ribosomes without side effects to the host’s ribosomes [13]. Antimicrobial agents benefit from ribosome subunit differences in selectively preventing bacterial growth. For instance, chloramphenicol binds to the 50S LSU while macrolides, aminoglycosides and tetracyclines bind to the 30S SSU [16]. Resistance to all major

classes of antibiotics covers the full array of mechanisms from target site alteration to antibiotic destruction [21].

Inhibitors of DNA replication. DNA is the main genetic material in organisms. DNA replication is important for cell viability and preventing the synthesis of any essential protein would be deleterious for cell survival [22]. DNA replication takes place in three enzymatically mediated steps of assembly: initiation, elongation and termination. The inactivation of a protein or enzyme involved in this process could inhibit the replication. DnaA is a protein that initiates replication by unwinding DNA in bacteria, which leads to elongation. Blocking the synthesis of DnaA protein (autoregulation) prevents DNA replication automatically. DnaA protein is a component that presents an obvious target to inhibit DNA replication in most bacteria [22]. On the other hand, DNA gyrase is an enzyme that can be targeted for inactivation [22,23]. This enzyme works to form supercoils in the elongation process.

Folic acid metabolism inhibitors. PABA or p-aminobenzoic acid is a coenzyme in the reactions of purine, pyrimidine and amino acid synthesis that involves adenosine triphosphate (ATP)-dependent condensation of pteridine to yield dihydropteroate acid, which then is converted to folic acid [13,14]. PABA, a substrate for folic acid catabolism, can be inhibited by drugs through a competitive inhibition mechanism [14].

Microbes possess defence mechanisms to survive attacks from foreign substances by means of antibiotic resistance genes. Resistance mechanisms operate to prevent harmful substances from entering the cell by modifying the membrane's access sites. If an antimicrobial substance manages to enter the cell, the bacteria have another survival strategy by using pumps on the cell walls to remove the substance. If these two methods fail to defeat the substance, the bacteria have certain enzymes that can prevent the substance from chemically bonding or modify the receptors targeted by the substance. The resistance gene can spread among the bacterial population through transduction via phages, conjugations between bacteria, or

transformation when a resistant bacterium releases its resistance genes and other bacteria pick those genes up.

Investigation of novel strains is one way to find a solution for the problem of antimicrobial resistance. A novel strain has a certain chance to carry a new compound that offers a different mechanism to destroy pathogens [24]. The bacterial domain contains many species with distinct characteristics. These characteristic differences between bacteria include differences in morphology and also their motility patterns on agar plates. One motility pattern that particularly deserves further observation is gliding.

Gliding bacteria live attached to solid surfaces and can move even on a dry medium. They can live in terrestrial or aquatic environments. Soil, decaying plant material, dung and bark are their typical habitats. However, they can also be found in sewage plants, algae, the surfaces of animals, on the exoskeletons of dead crustaceans, sand, mud, on the filling material of trickling filters, in the flakes of activated sludge, suspended in bodies of water, mouth and intestinal tracts of warm-blooded animals, eggs of water beetle, etc. The bacteria's gliding motility is the uniting character and may have the same taxonomic importance and reliability as flagellation [20].

Since flagellated bacteria will also occasionally swarm, we may have to seek other classification criteria in the way colonies of gliding organisms spread. Swarms of multiple gliding bacteria move slowly or only during their growth phase [20]. Gliding bacteria can also spread on rather dry agar surfaces. The swarm's edge usually shows flame-like protrusions, often bending back towards the swarm over considerable distances. The swarm surface itself is often structured, showing circular or radial ridges, knob-like accumulations of cells, or in the case of myxobacteria, dense fields of tiny migrating waves and fruiting bodies [20].

Many gliding bacteria's swarming movement occurs slowly or even only during their growth phase [20]. Swarming is the ability of colonies of gliding bacteria to spread over the surface of an agar medium while remaining physically thin and often veil-like [20]. The composition of the media, such as yeast agar (e.g. VY/2-

agar), water-agar and media with low peptone concentrations (0.1% and below) can stimulate the gliding motility and spread of the colonies [20]. Genetic analysis has revealed that, at least in myxobacteria, gliding motility is an extremely complex phenomenon [20].

Reichenbach and Dworkin (1981) stated that even though gliding bacteria are an artificial grouping, all or most gliding bacteria share the following characteristics: (1) All of them are Gram-negative and possess a typical Gram-negative cell wall; (2) All gliding bacteria are rod-shaped, and many of them form long, multicellular filaments; (3) No known gliding bacterium bears flagella at any time; (4) All gliding bacteria seem to produce slime, and moving cells leave a slime track behind them; (5) All the gliding bacteria that have been examined are more sensitive to actinomycin D than are other Gram-negative bacteria; (6) The FA spectrum of gliding bacteria seems to be dominated by odd-numbered and branched fatty acids, which is not typical for aerobic Gram-negative bacteria; (7) The quinones of the respiratory chain are exclusively menaquinones, which, again, is unusual for aerobic, Gram-negative bacteria; (8) Many of the gliding bacteria are yellow, greenish-yellow, orange, red and even violet [20].

Temperature and pH influence the optimum growth of bacteria. Bacteria can be divided into four different groups according to their preferred temperature range, namely psychrophilic (−5°C–15°C), mesophilic (30°C–37°C), thermophilic (50°C–60°C) and hyper-thermophilic (80°C–110°C) [13,25].

In terms of genomic nucleotide base content, bacteria have a G+C content within the range of 13–75 mol% [23]. The G+C content describes the prevalence of guanine and cytosine sites divided by DNA sequence length for bacterial genomes [23,26]. There are at least two orders, the *Myxobacterales* with their narrow G+C range (68–71 mol%) and the *Cytophagales* with their wide G+C range (28–67 mol%) [20].

This study focuses on the isolation of gliding bacteria, especially myxobacteria. The myxobacteria belong to delta-proteobacteria, Gram-negative organisms with an

outer membrane and unicellular bacteria with rod-shaped vegetative cells [20,21,27]. The vegetative cells assemble into a substantial mound and then configure a fruiting body (which consists of one to several sporangioles) by directed cell movement when the myxobacteria come under starvation conditions [21,28]. Cells/fruiting body colours vary from milky, yellow, orange, red, and brown to black [28]. Myxobacteria characteristics are gliding motility, a complex life cycle culminating in the formation of fruiting bodies and myxospores, the ability to degrade insoluble macromolecules and the production of an extensive battery of bioactive compounds [21].

Myxobacteria are mesophilic and grow well at 30°C. However, most myxobacterial strains can grow over a temperature range of 4°C–44°C [21,28]. Usually, vegetative cells cannot survive in temperatures above 45°C, but myxospores suspended in water tolerate 58°C–60°C [28]. Myxobacteria also need an environmental pH within the range of 6–8 to grow properly [21].

The compounds produced by myxobacteria are generally polyketides, non-ribosomal polypeptides and their hybrids, terpenoids, phenylpropanoids, and alkaloids [28]. Many of these substances show promising activity against bacteria, viruses, fungi, cancer cells, immune cells and malaria, exhibiting unusual action modes [28].

The myxobacteria contain a variety of bioactive compounds. Among over 2000 bacteriolytic strains of myxobacteria, 55% produce bioactive compounds; among over 700 cellulolytic myxobacteria, 95% produce bioactive compounds [21]. Gene clusters of myxobacteria produce some potential secondary metabolites; each strain in the same suborder has a possibility for producing the same compounds. Some of the secondary metabolites are listed in Table 2.1.

Many strains yield metabolites belonging to multiple structural classes and several chemical variants on each basic scaffold [28]. *Stigmatella aurantica* is a multiproducer of secondary metabolites, e.g. the Myxalamids, Stigmatellins, Aurachins, Myxochelins, Myxothiazols, and Myxochelins [29]. An unknown

siderophore from *Stigmatella aurantiaca* Sga15 led to a more detailed analysis of the fermentation broth and the identification of Myxochelin A that was isolated previously from other myxobacteria [30]. The Myxochelins are catecholate-type siderophores produced by some myxobacterial strains [31].

Sorangium cellulosum (suborder *Sorangineae)* produces Ambuticin VS, and Epothilon [29,32]. The Soce1525 strain, for instance, is a multiproducer and produce Chivosazoles, Sorangicin, Soraphens and Sulfangolide [33]. One *Sorangium cellulosum* strain also synthesises Sorangicin, Disorazol, Chivosazol and Sulfangolid, while another makes Disorazol, Icumazol and Soraphen [34].

Table 2.1 A sample of Myxobacteria strains and their secondary metabolites

Myxobacteria strain	**Secondary metabolites**
Stigmatella aurantica	Myxothiazol [32], Stigmatellin [29,32], Argyrin [32], Myxalamid [29], Myxochelin [29,31], Aurachin [29], Stigmolone27 [33,34], DKxanthene [35], Myxochromide [36], etc.
Archangium gephyra	Argyrin [37], etc.
***Myxococcus* strains**	Mixovirescin [29], Myxalamid A [36], Myxalamid B [36], Myxalamid C [36], Dkxanthen-534 [36], Dkxanthen-518 [36], Dkxanthen-560 [36], Dkxanthene13 [34], Myxochromid [36], Myxochromid-A2 [36], Myxochelin A [30,36,38], Myxochelin B [36,38], Cittilin-A [36], Myxothiazoles [35], Pseudochelin A [38], etc
Sorangium cellulosum	Ambruticin [28], Ambuticin [32], Epothilon [29,37], Myxochelin [29,31], Chivosazole [33,34], Sorangicin [33,34], Soraphen [33,34], Sulfangolide [33], Disorazole [34], Icumazol [34], etc.
Byssovorax Cruenta	Cruentaren A [39], etc.
Chondromyces	Cruentaren [39], Apicularen A [39], etc.

Epothilone, Tubulysin and Chondramide are observed only rarely among compounds of bacterial origin [37]. Epothilone is active against various tumour cell

lines, including breast carcinoma, HeLa cervical carcinoma, Burkitt's lymphoma, colon carcinoma, ovarian carcinoma, and neuroblastoma [37]. As the first derivative of Epothilone, Ixabepiloneas approved by the United States' Food and Drug Administration as a drug for breast cancer treatment [37]. Epothilone acts as a microtubule stabiliser, while Tubulysin acts as a polymerisation inhibitor of tubulin and shows antiangiogenic activity [37].

Some representative derivatives from five secondary metabolite classes known from *Myxococcus xanthus* identified by target screening in extracts from nine novel *Myxococcus* strains are Myxalamid A, Myxochelin B, DKxanthen-534, Myxochromid A2 and Citilin A [36]. Furthermore, Myxovirescin is produced by *Myxococcus xanthus* and *Myxococcus virescens* [29,33]. DKxanthene13 is found in the developmental cycles of *Myxococcus xanthus* [34].

Myxococcus and *Stigmatella* produce DKxanthenes as a class of yellow secondary metabolites [35]. One strain of *Myxococcus fulvus* can produce Myxothiazoles and up to 30 distinct derivatives, while a different strain of *Myxococcus fulvus* could produce only two derivatives of Myxothiazoles [34].

Within the myxobacteria's cellular life cycle, they utilise catechol, Myxochelin A and B to maintain their iron homeostasis [38]. Another product related to cell life cycles, Argyrin, was isolated from *Archangium gephyra* as a potent immunosuppressant that targets T-cell independent antibodies [37]. Species of the myxobacterial genus *Chondromyces* produce Cruentarens closely related to the class of Benzolactonenamides, including the Salicylihalamides and the Lobatamides. Also, Apicularen A, produced by several species of the myxobacterial genus *Chondromyces,* has structural similarities to that class [39].

Bacteria producing secondary metabolites are an important source of natural products with highly diverse structures and biological activities [36]. The development of methods for the analysis of secondary metabolites is still a matter of tremendous interest in discovering potentially new natural products [36]. Mass spectrometry–coupled liquid chromatography in conjunction with computational

and statistical tools can effectively capture microbial metabolic diversity with sensitivity and accuracy, thus enabling the targeted analysis and exploration of information-rich LC-MS datasets [36].

2.2 Hypothesis

As a large country with a tropical climate and a diverse geography that includes rainforests, seas and many mountain chains, Indonesia offers great potential for myxobacteria exploration. Lab cultivation results indicate that only around 1% of the naturally occurring bacterial community has been discovered and characterised so far [28]. Brockman investigated a moderate terrestrial habitat in 1976 and isolated strains of *Archangium, Chondromyces, Cystobacter, Myxococcus, Polyangium, and Stigmatella*. He reported a greater species diversity from regions with higher annual rainfall [28].

2.3 State of the art

Reichenbach (2005) reported that *Sorangium nigrum* strain Soce1654 was isolated in 1999 out of a sample from Bali, Indonesia [40]. Morgan (2010) noted that a *Myxococcus xanthus* sample isolated from Sulawesi (SUL2) had the worst performance in swarming against its *Arthrobactor globiformus* prey compared to other samples [41]. Garcia (2011) succeeded in isolating two novel strains of myxobacteria designated as SBSr002^T and SBSr003^T in 2007 from an Indonesian soil sample containing pieces of roots and other decaying plant material taken from the Landsweiler-Reden collection, Germany [42]. SBSr002^T (= DSM 24601^T = NCCB 100377^T), isolated in November 2007, and SBSr003^T (= DSM 24628^T = NCCB 100378^T), isolated in December 2007, were identified as a novel genus of *Polyangeaceae*, namely *Aetherobacter* [42].

Meanwhile, Meliana in 2018 reported that 10 myxobacteria (*Myxococcus fulvus, Myxococcus stipitatus* and *Melittangium lichenicola*) were successfully isolated from the Indonesian islands of Sumba and Papua [43]. Later in 2020, Meliana reported that a sample collected in Karimun and Simeuleu islands, Indonesia,

yielded 20 myxobacteria that belong to the genera *Myxococcus, Archangium* and *Corallococcus* [44].

2.4 Materials and methods

2.4.1 Materials

The GINAICO team collected the samples in 2016 and 2017. GINAICO stands for the German-Indonesian Antiinfective Cooperation, which works under the aegis of the DAAD and LIPI. This study was funded by The Federal Ministry for Education and Research.

Soil, decayed wood and coral rocks were sample sources for this research. They were collected from Indonesia, specifically Mt. Padang, the Bogor Botanical Garden and Mt. Tangkuban Perahu, all in West Java. Samples were also taken from Pandawa Beach in Bali, around Malang, East Java and Mangrove swamps in Sulawesi.

2.4.2 Methods

2.4.2.1 Isolation of the bacteria

Screening. The samples were dried at room temperature (25–30°C). An approximately one-gram sample was poured onto the medium for the isolation of myxobacteria. Two media target different organisms for isolation: a water-agar medium for predators and ST21-agar for cellulose decomposers. The water-agar medium consisted of 0.15% $CaCl_2.H_2O$, 0.15% $MgSO4.7H_2O$, 50 mM HEPES, 1 $mL.L^{-1}$ vitamin solution and 1.8% agar. The composition of the vitamin solution was 0.2% biotin, 2.0% nicotinic acid, 1.0% thiamine, 1.0% 4-aminobenzoic acid, 0.5% pantothenic, 5.0% pyridoxamine, and 2.0% cyanocobalamin. The vitamin solution was sterilised by filtration and stored at 4°C. On the other hand, the medium for cellulose decomposers starts with solution A: 0.1% K_2HPO_4, 0.002% yeast extract or 1% baker's yeast, and 1.0% agar. This was dissolved in two-thirds of the eventual total volume of distilled water. Then solution B, consisting of 0.1% KNO_3, 0.1% $MgSO_4.7H_2O$, 0,1% $CaCl_2.2H_2O$, 0.002% $FeCl_3$, 0.01%

$MnSO_4.7H_2O$, was dissolved in the remaining third of the total volume. The two solutions were then autoclaved separately. Afterward, solutions A and B were combined with a 1 $mL.L^{-1}$ trace element solution into the combination medium. Supplement 1 presents the composition of the trace element solution. The autoclaved *Escherichia coli* was streaked upon the water-agar media prior to the sample. Meanwhile, a sterile paper filter with 1x1 cm dimension was placed upon the ST-21 medium.

Isolation. The observation of the plate during the incubation period started on the fourth day. This gave time for the growth of fruiting bodies and the emergence of special behaviour from myxobacteria, such as swarming and gliding. The fruiting body or a part of agar that contains swarming or gliding bacteria was removed using a sterile injection needle. This was then transferred to a new water-agar plate for the next process: purification.

Purification was a combination of transfers from one agar medium to another and observation under the microscope until a pure isolate has been produced. First, the object for purification and its surrounding area was cut off from the first plate. Then it was transferred to a new agar plate. Several different agar media were used for purification in this research. First was water-agar, which contains cyclohexamide, soraphen, and vitamin. The second was water-agar with levamisole, cyclohexamide, soraphen, and vitamin. The third was water-agar, P agar, Cy agar, and VY/2 agar containing vitamin B12, respectively. After purification, a square area of agar from the last media of purification was cut off and placed in 1 to 20 mL Cy/H medium. It was then put in the shaker incubator at 30°C and 160 rpm for 5 days prior to being transferred to a 100 mL solution. The decision on whether we had achieved a purified isolate was based on microscopic observation using an Olympus Stereomicroscope SZX 10 and a Zeiss Standard Microscope. Supplement 1 presents the composition for all media.

2.4.2.2 Extraction, polymerase chain reaction and sequencing of DNA

2.4.2.2.a Extraction of genomic DNA

Firstly, the cell wall had to be lysed in order to obtain genomic DNA from the target of the analysis. We used the protocols described in the Invisorb® Spin Plan Mini Kit (250) manual, Stratec Molecular, REF 1037100300 to perform DNA isolation. For this, we needed a 1 mL cell mass solution to put into a 2 mL tube. Precipitation of the cell mass was accomplished through centrifugation at 10000 rpm for 5 min. The sediment was then separated and used for the next step. 100 µL of lysis buffer P was pipetted to the sediment, and the mixture was taken through a brief vortex. Then it was shaken at 95°C with a thermomixer for 1 min. 300 µL of lysis buffer P and 20 µL proteinase K were added; then, the whole mixture was homogenised with the pistil and incubated at 65°C for 30 min.

The next step is **filtration** to separate the DNA from the cell-residue. A pre-filter was inserted into a 1.5 mL tube. The lysis solution was transferred onto the pre-filter and centrifuged at 11000 for 1 min, after which the pre-filter was discarded.

After that came the **binding step**. 200 µL of binding buffer A was added to the filtrate in the tube and vortexed thoroughly. The suspension was then transferred onto a spin filter in a new 1.5 mL tube. It was incubated for 1 min at room temperature, then centrifuged at 10000 rpm for 2 min. After that, the filtrate was discarded, and the spin filter was placed inside the tube once more. This is the end of the binding step.

The next step is **washing**. It seeks to rinse the impurities from the bound DNA. 550 µL wash buffer I was added to the spin filter, which was then centrifuged at 11000 rpm for 1 min. The filtrate was discarded, and the spin filter was put back into the tube. The washing step was then repeated with washing buffer II. The filtrate was once again discarded, and the spin filter was put back into the tube and centrifuged for 4 min at 11000 rpm to complete the drying process.

After the washing step, **elution** processed. This process sought to obtain clean genomic DNA by eluting the column. The spin filter from the previous step was

transferred onto a new 1.5 mL tube. 30 µL of pre-warmed elution buffer was added and incubated for 3 min at room temperature. Then it was centrifuged at 11000 rpm for 1 min. Afterwards, the spin filter was discarded, and the DNA stored at 4°C.

2.4.2.2.b Gel electrophoresis

Gel electrophoresis is a method to evaluate the quality of the DNA that has been previously isolated based on the appearance of the fragments. It also identifies the fragments' approximate size compared to a reference. We used a gel of 0.8% agarose [27] (Broth Art-Nr. 3810.3) in 1x tris-acetate-EDTA (TAE) buffer.

We prepared 40 mL of the agarose according to the instructions on the label. It was then run at 70 volts and 400 mA for 45 min utilising BIORAD Power PAC 300. The gel was observed with a Herolab Transluminator B-1393-307N, serial no. TCH-KBL 14406009, 14 MP (205) with E.A.S.Y Win32 version 6.1.

2.4.2.2.c Polymerase chain reaction

Polymerase chain reaction (PCR) is a method to perform DNA replication. The method required Taq DNA polymerase, forward primer, reverse primer, template DNA, and water PCR. We used the JumpStart™ Taq ReadyMix™ from SIGMA-ALDRICH with catalog number P2893 as the Taq DNA polymerase.

We also used PCR water from Roth Art-Nr. T143.3. In addition, we used universal primers from Eurofins Genomic such as 27F (5′-AGAGTTTGATCMTGGCTCAG-3′), and 1492R (5′-TACGGYTACCTTGTTACGACT-3′) or 1525R (5′-AAGGAGGTGWTCCARCC-3′). According to International Union of Pure and Applied Chemistry (IUPAC) nucleotide names, M is A or C, Y is C or T, and W is A or T.

2.4.2.2.d Strain identification

We sent the DNA sample to the sequencing center in Genome Analytics-GMAK HZI. The primers for this analysis were 27F (5′-AGAGTTTGATCMTGGCTCAG-3′), 518R (5′-CGTATTACCGCGGCTGCTGG-3′), 1100F (5′-YAACGAGCGCAACC-3′), and 1100R (5′-GCGTTGCGCTCGTTG-3′). Meanwhile, the sequential analysis was performed with BioEdit and BLAST.

The sequencing reaction was performed on a 10 µL scale using the BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific), 10 pmol of primers and 500 ng of template DNA. The following thermal cycle was used for the amplifications: 96°C for 1 min, followed by 25 cycles of 96°C for 10 s, 60°C for 4 min. Excess dye terminators were removed using ethanol precipitation. The samples were resuspended in 10 µL Hi-Di formamide and were electrophoresed on an Applied Biosystems 3730xl automated DNA sequencing instrument according to the manufacturer's instructions. The strains were then compared to assess their similarities to reference strains in the NCBI database.

2.4.2.3 Antiinfectives bioassay

The bioassay was used to assess the strains that we isolated during the study for specific activity against some bacterial and fungal pathogens from the MISG laboratory library. The strains' activity was analysed through serial dilution tests. This method required a methanol extract from the strain culture, test bacteria, growth media, 99.5% methanol as a negative control, and an antibiotic as a positive control.

2.4.2.3.a Sample preparation

The methanol extract was prepared from 100 mL medium that contained 2% XAD. It was fermented for five days at 30°C using a shaker incubator with agitation at 160 rpm. The filtrate was separated from XAD with a sieve. Then the XAD was washed with distilled water and dried. Afterward, it was soaked with ± 100 mL acetone for at least 1 hour, and the filtrate was separated from the XAD with filter paper.

The filtrate was then evaporated with a Heidolph Laborota 4003 rotavorator at a controlled temperature of 40°C until it was dry. Finally, the extract was dissolved in 1 mL of methanol and transferred to a 1.5 mL tube. The extracts were kept at −20°C until they were put to use. We prepared 10 kinds of media for use, namely A, Cy, Cy/H, E, H, Myxovirescin, P, Pol, S, and VY/2. Supplement 1 explains the compositions of these media.

2.4.2.3.b Test strains preparation

The test strains were representative of Gram-negative bacteria, Gram-positive bacteria, and fungi. We took *Escherichia coli* DSM 1116 as the Gram-positive test bacteria for this study, *Staphylococcus aureus* Newman as the Gram-negative bacteria, and *Candida albicans* DSM 1665 as the yeast. The bacteria were prepared in a Mueller-Hinton broth medium. The yeast was prepared in a Myc medium. Each bacterium had an optical density of 600 nm in its medium. Supplement 1 provides the relevant media compositions.

2.4.2.3.c Serial dilution test

The total volume in the first line of each well of 96's well-plate was 300 µL. It contained 20 µL of sample and 280 µL of test strain solution. Starting from the second to the eighth line, there was 150 µL in volume of test bacteria. The dilution began from the first line and proceeded to the eighth line until the total volume in each well was 150 µL. The plates were arranged with the sample and positive and negative controls side-by-side. Methanol was used as a negative control, while Gentamicin and Nystatin were used as positive controls for the bacteria and yeast, respectively. Incubation was carried out at 30°C for 18–24 hours. The observation was continued for 48 hours to control the result. The result was recorded as being positive when the bacteria could not grow in the well.

2.4.3 Identification of active compounds

The compounds from the extracts were detected by HPLC (High Performance Liquid Chromatography) and High-Resolution Electrospray Ionisation Mass Spectrometry (HRESIMS) methods. Retention time and mass were the main parameters in analysing the compounds. MyxoBase and the Dictionary of Natural Products were used as databases to identify the compounds' novelty.

2.5 Result and discussion

2.5.1 Isolation of myxobacteria

We had isolated 59 gliding bacteria samples from Indonesia for this project in 2016–2017. According to the 16S rRNA gene sequencing results, 98% of the isolates can be categorised as myxobacteria, while one is a gliding bacterium from the phylum *Bacteriodetes*. Table 2.2 lists the names of all the strains.

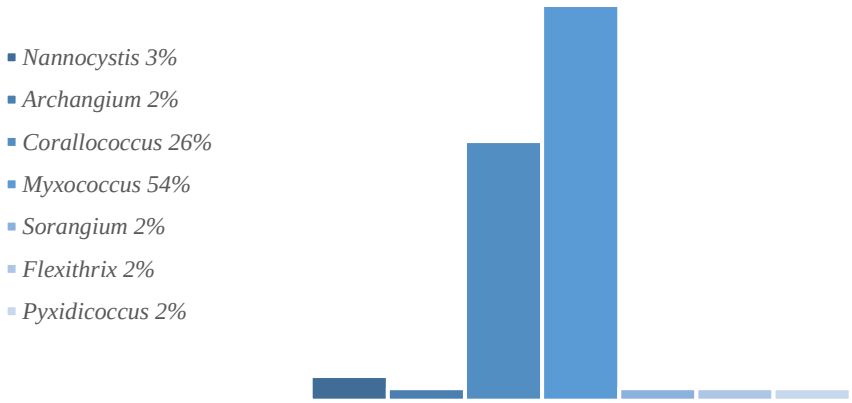


Chart 2.1 Percentage of strains of gliding bacteria from Indonesia's biodiversity

The data on Chart 2.1 shows that the largest proportion of gliding bacteria isolated from the Indonesian samples are from the suborder *Cystobacterineae*, especially the family *Myxococcaceae*. The suborder *Cystobacterineae* in order *Myxococcales* comprise four genera validly described as *Myxococcus*, *Corallococcus*, *Pyxidicoccus* and *Aggregicoccus* [28,45].

The results of isolates in this project were similar to those reported by Reichenbach (2005) in that the most frequently encountered genera were *Myxococcus* and *Corallococcus* [21]. *Myxococcus* and *Corallococcus* turned out to be the most commonly isolated myxobacteria genera in samples of soil, tree bark, rotting wood, freshwater and marine environments [45,46]. Picking the right climate seems to be important in obtaining the desired myxobacteria distribution [21]. On the other hand, it is just as important to improve isolation and purification procedures to

accommodate the specific and often unusual requirements of novel and rare myxobacteria [37].

Myxococcus xanthus has become the primary and best-studied object of myxobacterial research [46]. In addition to *Myxococcus*, *Sorangium cellulosum* represents one of the rare groups of aerobic Gram-negative bacteria that degrade cellulose for use as a carbon source [47]. Myxobacteria are generally considered mesophilic soil microbes with an optimum living temperature of 30°C and as bacteria that synthesise many active compounds [46].

One of the isolated strains is Soce1964KM, which was identified by the 16S rRNA gene sequence analysis as a *Sorangium*. It has a 100% similarity with *Sorangium sp*. 321 (Soce321^{T}) with accession number MG 824980.1. Mohr et al. (2018) isolated the Soce321^{T} (DSM 53339^{T}, NCCB 100640^{T}) sample from Bulgarian soil. It has a G+C content of 71.4 mol% a [48]. Soce1964KM fruiting bodies are coloured orange to brownish. It grows at 30°C in a Cy/H medium adjusted with vitamin B12. Soce1964KM needs 14 days to grow in liquid culture. Its growth was observed to be more stable in VY/2 containing vitamin B12.

Myxospores and fruiting bodies are desiccation resistant and survive in air-dried soil at environmental temperatures for long periods. *Sorangium cellulosum*, *Corallococcus*, *Nannocystis*, *Archangium*, and other myxobacteria have been successfully isolated from the same air-dried sample after 12 years of storage. Myxobacteria have never been observed to be pathogenic towards higher plants and animals. [40]. Table 2.2 lists the gliding bacteria isolated from Indonesian environments in this project. Meanwhile, Figure 2.1 presents several photographs of gliding bacteria found in this study.

Table 2.2 Strains isolated from Indonesia's microbial biodiversity and their similarity compared to type strains

No.	Code Strain	Sample Source	Similarity (%)	Closest Type Strain	Accession Number of Type Strains
1.	1900-2 KM	Sulawesi	99.64	*Nannocystis pusilla* Na p29 ^{T}	NR104789.1
2.	NaFK3KM	Sulawesi	99.75	*Nannocystis pusilla* Na p29^{T}	NR104789.1
3.	1963-2 KR	Bali	98.98	*Archangium gephyra* DSM 2261 ^{T}	CP011509.1
4.	CCMK6KM	Sulawesi	99.67	*Corallococcus coralloides* DSM 2259 ^{T}	CP003389.1
5.	CCMK5KM	Sulawesi	97.30	*Corallococcus coralloides* DSM 2259 ^{T}	CP003389.1
6.	MyxoGP241TM	West Java	99.23	*Corallococcus coralloides* DSM 2259 ^{T}	CP003389.1
7.	CeGP13TM	West Java	99.29	*Corallococcus coralloides* DSM 2259 ^{T}	CP003389.1
8.	MyxoGP11TM	West Java	99.72	*Corallococcus exiguus* DSM 14696 ^{T}	NR115864.1
9.	CeGP62TM	West Java	99.72	*Corallococcus coralloides* DSM 2259 ^{T}	CP003389.1
10.	CeGP11TM	West Java	99.40	*Corallococcus coralloides* DSM 2259 ^{T}	CP003389.1
11.	MxGP61TM	West Java	99.59	*Corallococcus coralloides* DSM 2259 ^{T}	CP003389.1
12.	MxGP62TM	West Java	99.66	*Corallococcus coralloides* DSM 2259 ^{T}	CP003389.1
13.	MxGP63KM	West Java	99.66	*Corallococcus coralloides* DSM 2259 ^{T}	CP003389.1
14.	CCGP6KM	West Java	99.67	*Corallococcus coralloides* DSM 2259 ^{T}	CP003389.1

15.	Cce1969-1KR	Bali	99.89	*Corallococcus coralloides* DSM 2259 **T**	CP003389.1
16.	1938-1KR	Bali	100	*Corallococcus coralloides* DSM 2259 T	CP003389.1
17.	BGB2137RAWA	West Java	99.89	*Corallococcus coralloides* DSM 2259 T	CP003389.1
18.	Ce2131KM	West Java	99.72	*Corallococcus coralloides* DSM 2259 T	CP003389.1
19.	CC2132KM	West Java	99.78	*Corallococcus coralloides* DSM 2259 T	CP003389.1
20.	CC2137KM	West Java	99.89	*Corallococcus coralloides* DSM 2259 T	CP003389.1
21.	Ce2146KM	West Java	99.78	*Corallococcus coralloides* DSM 2259 T	CP003389.1
22.	BGB2139RAST02	West Java	99.64	*Corallococcus coralloides* DSM 2259 T	CP003389.1
23.	3RB1-1	West Java	100	*Corallococcus exiguus* DSM 14696 T	NR115864.1
24.	Soce1969-2KR	West Java	99.73	*Corallococcus coralloides* DSM 2259 T	CP003389.1
25.	MxGP53TM	West Java	99.2	*Myxococcus stipitatus* DSM 14675 T	CP004025.1
26.	MyxoGP31TM	West Java	99.33	*Myxococcus stipitatus* DSM 14675 T	CP004025.1
27.	CeGP512TM	West Java	99.51	*Myxococcus stipitatus* DSM 14675 T	CP004025.1
28.	MxGP43TM	West Java	99.20	*Myxococcus fulvus* 124B02 T	CP004025.1
29.	MxsGP3KM	West Java	99.55	*Myxococcus stipitatus* DSM 14675 T	CP004025.1
30.	MxGP1KM	West Java	99.63	*Myxococcus stipitatus* DSM 14675 T	CP004025.1
31.	MxGP5KM	West Java	99.57	*Myxococcus stipitatus* DSM 14675 T	CP004025.1

32.	BGB2139RAST03	West Java	99.79	*Myxococcus stipitatus* DSM 14675 **T**	CP004025.1
33.	Mx2134KM	West Java	99.66	*Myxococcus stipitatus* DSM 14675 T	CP004025.1
34.	Mx2132KM	West Java	100	*Myxococcus stipitatus* DSM 14675 T	CP004025.1
35.	MxGP4KM	West Java	99.41	*Myxococcus stipitatus* DSM 14675 T	CP004025.1
36.	CeGP43TM	West Java	99.63	*Myxococcus stipitatus* DSM 14675 T	CP004025.1
37.	MxMK7KM	Sulawesi	99.60	*Myxococcus macrosporus* DSM 14697 T	CP022203.1
38.	MxMK5-1KM	Sulawesi	99.76	*Myxococcus macrosporus* DSM 14697 T	CP022203.1
39.	MxGC9KM	West Java	100	*Myxococcus macrosporus* DSM 14697 T	CP022203.1
40.	Mx1941KM	Bali	99.88	*Myxococcus macrosporus* DSM 14697 T	CP022203.1
41.	MxGP1962KM	Bali	99.52	*Myxococcus macrosporus* DSM 14697 T	CP022203.1
42.	Mxx1929(1) KR	Bali	99.73	*Myxococcus macrosporus* DSM 14697 T	CP022203.1
43.	Mxx1929KR	Bali	99.60	*Myxococcus macrosporus* DSM 14697 T	CP022203.1
44.	Mx1955KM	Bali	99.52	*Myxococcus macrosporus* DSM 14697 T	CP022203.1
45.	TP13RAWA-A	West Java	98.93	*Corallococcus exiguus* DSM 14696 T	NR115864.1
46.	TP13RAWA-B	West Java	99.76	*Myxococcus macrosporus* DSM 14697 T	CP022203.1
47.	TP13RAWA-C	West Java	99.64	*Myxococcus macrosporus* DSM 14697 T	CP022203.1

48.	Mx1960 KM	Bali	99.60	*Myxococcus macrosporus* DSM 14697 **T**	CP022203.1
49.	Mx1960KR (1)	Bali	99.67	*Myxococcus macrosporus* DSM 14697 T	CP022203.1
50.	Mx2133KM	Bogor	99.87	*Myxococcus macrosporus* DSM 14697 T	CP022203.1
51.	MxGP1941KM	Bali	99.88	*Myxococcus macrosporus* DSM 14697 T	CP022203.1
52.	Mx2129KM	Jakarta	100	*Myxococcus macrosporus* DSM 14697 T	CP022203.1
53.	MxGP1963KM	Bali	99.73	*Myxococcus macrosporus* DSM 14697 T	CP022203.1
54.	Mx1956KM	Bali	99.66	*Myxococcus macrosporus* DSM 14697 T	CP022203.1
55.	Mx1965KR	Bali	99.93	*Myxococcus fulvus* ATCC 25199 T	NR043946.1
56.	Mx1809KM	Jakarta	99.88	*Myxococcus fulvus* ATCC 25199 T	NR043946.1
57.	Soce1964KM	West Java	100	*Sorangium* sp. Soce321T	MG824980.1
58.	1932KM	Bali	90.51	*Flexithrix dorotheae* ATCC 23163T	NR112649.1
59.	A1932RA b1a1	Bali	99.62	*Pyxidicoccus fallax strain* DSM 14698T	NR043948.1

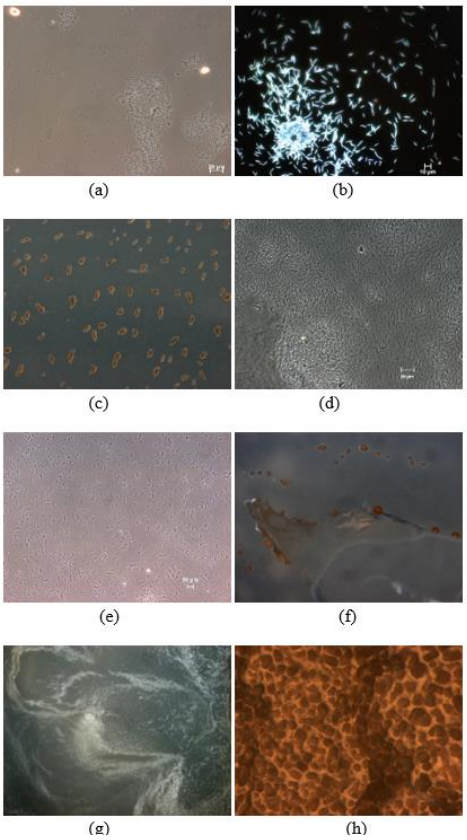


Figure 2.1 Representatives of gliding bacteria isolated from Indonesian biomes. (a) 1932KM; *Flexithrix*, (b) 1963-2KR; *Archaengium*, (c) 1938-1; *Corallococcus*, (d) A1932RA-b1a1; *Pyxidicoccus*, (e) MxGP53TM; *Myxococcus*, (f) MyxoGP31TM; Myxococcus, (g) NaFK3KM; *Nannocystis*, and (h) Soce1964KM; *Sorangium*

2.5.2 Bioactivity screening and secondary metabolite analysis on gliding bacteria samples from Indonesia

Antibiotic resistance is a global issue that affects all countries and motivates never-ending research to find new antibiotics. This issue of antibiotic resistance drives the scientific interest in gliding bacteria in the hopes of finding candidate antibiotic compounds. *Myxobacteria*'s status as non-pathogenic bacteria offers certain advantages in the pursuit of this study's goals.

The search for antibiotic resistance solutions involves looking for new antibiotic compounds and/or new mechanisms of action against pathogens. *Escherichia coli*, *Staphylococcus aureus*, and *Candida albicans* were chosen as the pathogens against which the 59 gliding bacterial isolates from Indonesia were tested with 10 methanol extracts.

The methanol extracts of some strains showed activity against the test bacteria in the fermentation media, such as A, C, Cy/H, E, H, P, Pol, Myxovirescin, S, and VY/2. Soce1964KM, which belongs to the genus *Sorangium* as a rare myxobacterium, inhibited *Staphylococcus aureus'* growth in five out of ten fermentation media and all the results displayed the highest level (up to H) in the 96 well plates. The sample in the methanol extract of E medium showed positive activity not only against *Staphylococcus aureus* but also against *Candida albicans*.

Additionally, the *Myxococcus* and *Corallococcus* strains MyxoGP53TM, MyxoGP43TM, MxsGP3KM, and CeGP43TM showed positive results against *Staphylococcus aureus* and *Candida albicans* in the VY/2 medium. Meanwhile, MxGP5KM displayed the ability to inhibit *Candida albicans*, while MxGC9KM showed antibacterial activity against *Escherichia coli*. Chart 2.2 lists the results.

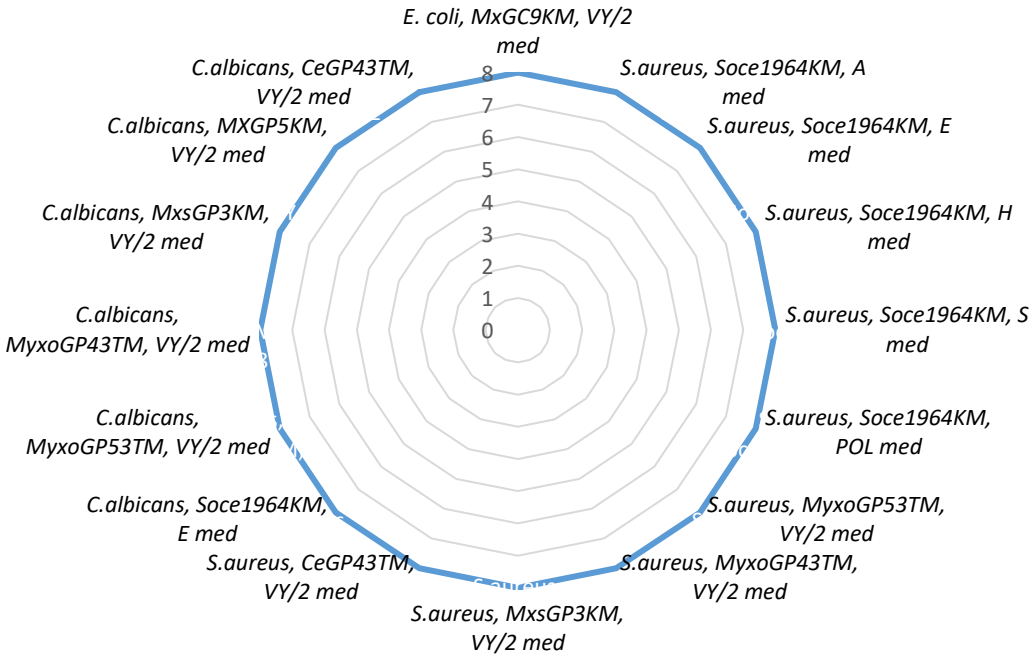


Chart 2.2 Screening profile of the best antiinfective results from Myxobacteria strains against *Escherichia coli*, *Staphylococcus aureus*, and *Candida albicans*. A score of 8 is equal to an activity level of H on the 96 well plates; MxGC9KM, Soce1964KM, MyxoGP53TM, MyxoGP43TM, MxsGP3KM, MxGP5KM, and CeGP43TM = isolated strain's name; A, E, H, S Pol and VY/2 = fermentation media

Compound analysis performed upon the methanol extract of fermentation results from *Myxococcus* and *Corallococcus* through HPLC and HRESIMS showed that the extracts contain Myxovirescin, Myxalamids, Myxochelin B, and DKxanthene. By contrast, the extracts obtained from Soce1964KM (of the family *Sorangineae*) contained Disorazol.

These compounds' mechanisms of action have already been discovered in prior studies. For instance, Myxovirescins were discovered by Rosenberg et al. in *Myxococcus xanthus* [49]. Myxovirescins are excreted during the late exponential and early stationary growth phases and display strong inhibitory activity against growing bacterial cells, even when applied at concentrations of less than 5 $\mu g.mL^{-1}$ [49]. The low concentration needed for bacterial inhibition activity showed that the secondary metabolites are important [13]. The Myxovirescins comprise a family of closely related antibiotics featuring a distinctive 28-membered macrolide ring [49]. The antibiotic interferes with cell wall biosynthesis by inhibiting a novel target, i.e. the type II signal peptidase LspA, which is involved in the maturation of lipoproteins required for murein biosynthesis [49].

Potent antimicrobial activities distinguish Myxovirescins and Myxalamids. Myxalamids are inhibitors of electron transport in the respiratory chain. They blocked the electron flow at complex I of the mitochondria (NADH:ubiquinone oxidoreductase) in a competitive manner, but do not act on bacterial complex I. This explains why Myxalamids are mainly active against fungi [49].

Myxochelin is a catechol siderophore produced by many myxobacteria [38]. They produce the substance to maintain their iron homeostasis since is indispensable for microbial viability. Myxochelin acts as a component of several enzymatic processes such as respiration, amino acid synthesis, oxygen transport and DNA biosynthesis. These molecules are secreted from the cells to solubilise and capture ferric iron. Subsequently, the ferric iron-siderophore complex is actively transported back into the bacterial cytoplasm, and the coordinated ion is released by either reductive or hydrolytic mechanisms [38]. In terms of pathogenesis, there is often competition

for iron between the pathogen and the host [50]. Indeed, siderophores have been implicated as virulence factors for several eukaryotic pathogens primarily because of their ability to remove bound iron from haemoglobin and other host proteins [50,51].

Meanwhile, DKxanthenes are a family of yellow pigments required to form viable spores in myxobacterial development, and these compounds were originally characterised by *Myxococcus xanthus* [52,53]. DKxanthenes are produced by yellow-phase cells but not by tan-phase cells [35]. The presence of such pigments in *Myxococcus xanthus* was suspected for decades, because growth phase cells give rise to bright yellow colonies (Burchard et al., 1977) [52].

We also identified a known Disorazol from the genus *Sorangium.* This Disorazol shows strong activity against our test strains. Disorazols, in general, display strong activity as tubulin depolymerising agents [53].

The derivatives of Disorazol (in patent number WO 2004/053065 A2) are DszA, DszB, DszC and DszD, which were obtained from Soce12 [54]. Meanwhile, in patent number WO 2006/07503 A1, the inventors had isolated Disorazols A, B, C, D, E, F, G, H, I and derivatives thereof [54]. Among these secondary metabolites, they claimed that Disorazoles and their derivatives display activity in inhibiting tubulin polymerisation, inducing apoptosis and arresting cell proliferation cycles at concentrations as low as e.g., 3 pM [54].

These active compounds are produced by a gene cluster responsible for producing the secondary metabolites through biosynthesis formation [13,29,31]. This gene cluster contains the kind of genes frequently associated with Polyketide Synthases (PKS) and Non-Ribosomal Peptide Synthases (NRPS) for the production of secondary metabolites in most bacteria and fungi [29,55]. Polyketide Synthases (PKS) and NRPS are large multifunctional and modular enzyme systems. They are responsible for the biosynthesis of an enormous diversity of biologically active natural products [56]. In myxobacteria, the PKS and/or NRPS use biosynthetic

proteins to catalyse the stepwise aggregation from short-chain carboxylic acids or from amino acids [33].

2.6 Conclusion

Samples from the Indonesian microbial biodiversity could yield novel or rare strains with or without an associated active compound. Through the GINAICO programme, 59 strains of the gliding bacteria have been isolated, including a rare myxobacteria, a member of the genus *Sorangium,* and a novel non-myxobacteria. Although the 59 myxobacteria in this study were known species, all the strains offer new information about the biodiversity of microbes in Indonesia. This is because gliding bacteria from Indonesian samples remains a relatively novel research topic. Thus, there is an opportunity for further studies on the subject, focusing on other sample resources. Meanwhile, the novel non-myxobacteria strain demands further study for polyphasic taxonomy analysis.

2.7 Summary

The isolation methods we used for myxobacteria succeeded in discovering gliding bacteria. Some of the isolated myxobacterial strains showed antibacterial activity, but unfortunately, further analysis showed that the associated active compounds are already known to the scientific community.

2.8 Recommendation

Finding new drugs or antimicrobials from bacteria is a challenging endeavour. It starts with isolating the strain, screening for bioactivity and identifying active compounds, all of which are complicated processes. This is because bacteria are a special category of a living organism whose metabolism depends heavily upon environmental conditions. When they are observed in media with different parameters (such as temperature, pH and nutrients) from their original environment, the difference alters their cellular processes enough to become impossible to obtain new secondary metabolites from them.

The critical points in each step of a strain's isolation should be known. Some recommendations on things that deserve particular attention in each step are described below:

1. Sampling phase. Every site carries the possibility of harbouring a rare or new strain, but extreme environments tend to have a higher probability for this. During sample collection, environmental conditions (such as date and time of sampling, local weather, temperature, pH from the sample's source and geographical coordinates) should be recorded in detail. These records would help the researchers replicate environmental circumstances that could enhance the bacteria's chances of surviving and thriving in the laboratory.
2. Isolation phase. The standard isolation medium for the target bacteria may have to be adapted to account for its origins. Some materials from the bacteria's original habitat should be added to the medium prepared in the laboratory. The assumption is that this might help the target bacteria grow well in otherwise unfamiliar conditions. For example, when we isolate a bacterium from a soil sample, then an extract of the soil in distilled water should be added to the bacterium's isolation medium. Once the pure isolate has been obtained, a backup should be prepared in cryo tubes and kept at −70°C.
3. Fermentation phase. Guidelines on media, optimum pH, temperature, incubation time and velocity of the shaker incubator are in the existing literature. But these parameters should be adapted to laboratory conditions. Adjustments for local environmental conditions can influence the bacteria's chances of survival. Adding an extract of their original habitat to their growth medium could decrease the stress they experience and increase the chance that they would produce secondary metabolites the same way they do in their original habitat.

In my opinion, before we start to isolate target bacteria from their extreme environments, it would be better to perform an analysis of the microbial community first using specific primers conserved from the natural habitat of the target bacteria. This way, we would observe the bacterial population in our sample and then

continue to isolate the target bacteria. Bioengineering can be used as a solution to enhance their productivity in this regard.

2.9 Acknowledgements

I would like to thank Kathrin I. Mohr and Tjandrawaty Mozef for their kindness in sharing their isolates for further analysis. Thank you to Birte Trunkwalter and Stephanie Schultz for their assistance in the isolation of myxobacteria. Also, I thank Enge Sudarman, Gian Primahana, Aileen Gollasch, Klaus Peter Conrad and Sabrina Karwehl for helping me in the chemical analysis.

2.10 References

[1] R. Cribb and M. Ford, "Indonesia as an archipelago: managing islands, managing the seas," https://ses.library.usyd.edu.au/bitstream/handle/2123/16146/Cribb%20&%20Ford%20Indonesia%20as%20an%20Archipelago.pdf;jsessionid=F5AD3D11C27DCD4590743940F7DF6707?sequence=2.

[2] R. W. Matindas, A. Poniman, S. Martha et al., *Atlas nasional Indonesia (Indonesian National Atlas)*: *Fisik dan lingkungan alam (Physical and natural environment)*, BAKOSURTANAL, Indonesia, 2008.

[3] W. B. Whitman, D. C. Coleman, and W. J. Wiebe, "Prokaryotes: the unseen majority," *Proceedings of the national academy of sciences*, vol. 95, pp. 6578–6583, 1998.

[4] Y. Chen, J. Kuang, P. Jia, et al., "Effect of environmental variation on estimating the bacterial species *Richness*," *Frontiers in microbiology*, vol. 8, p. 690, 2017.

[5] M. Kaspari, M. N. Garcia, K. E. Harms, et al., "Multiple nutrients limit litterfall and decomposition in a tropical forest," *Ecology letters*, vol. 11, no. 1, pp. 35–43, 2008.

[6] A. T. Nottingham, L. C. Hicks, A. J. Q. Ccahuana et al., "Nutrient limitations to bacterial and fungal growth during cellulose decomposition in tropical forest soils," *Biology and Fertility of Soils*, vol. 54, no. 2, pp. 219–228, 2018.

[7] F. Li, L. Chen, J. Zhang, et al., "Bacterial community structure after long-term organic and inorganic fertilization reveals important associations between soil nutrients and specific taxa involved in nutrient transformations," *Frontiers in microbiology*, vol. 8, p. 187, 2017.

[8] H. Smith, "Microbial surfaces in relation to pathogenicity," *Bacteriological reviews*, vol. 41, no. 2, pp. 475–500, 1977.

[9] W. Martin and E. V. Koonin, “A positive definition of prokaryote,” *Nature*, vol. 442, no.868, 2006.

[10] D. A. Walsh and W. F. Doolittle, “The real domains of life,” *Current biology*, vol. 15, pp. R237-240, 2005.

[11] B. Alberts, A. Johnson, J. Lewis, et al., *Molecular biology of the cell*, New York: Garland Science, 4th edition, 2002.

[12] G. Kapoor, S. Saigal, and A. Elongavan, “Action and resistance mechanisms of antibiotics: a guide for clinicians,” *Journal of anaesthesiology, clinical pharmacology*, vol. 33, no. 3, pp. 300–305, 2017.

[13] G. F. Brooks, S. A. Morse, K. C. Carrol, et al., *Medical microbiology*, The McGraw-Hill Companies, Inc., United States, 26th edition, 2013.

[14] S. Kırmusaoğlu, N. Gareayaghi, and B. S. Kocazeybek, “Introductory chapter: the action mechanisms of antibiotics and antibiotic resistance, antimicrobials, antibiotics resistance, antibiofilm strategies and activity methods,” *IntechOpen*, pp. 1-9, 2019.

[15] D. D. Nikolaeva, M. S. Gelfand, and S. K. Garushyants, “Simplification of ribosomes in bacteria with tiny genomes,” *bioRxiv* 755876, 2020.

[16] F. C. Tenover, “Mechanisms of antimicrobial resistance in bacteria,” *The American journal of medicine*, vol. 119, 6 Suppl 1, S3-10; discussion S62-70, 2006.

[17] G. E. Fox, “Origin and evolution of the ribosome,” *Cold spring harbor perspectives in biology*, vol. 2, no. 9, a003483, 2010.

[18] A. Filipovska and O. Rackham, “Specialization from synthesis: how ribosome diversity can customize protein function,” *FEBS letters*, vol. 587, no. 8, pp. 1189–1197, 2013.

[19] N. Ito, M. B. Anam, S. A. I. Ahmad, et al., “Transdifferentiation of human somatic cells by ribosome,” *Development, growth & differentiation*, vol. 60, no. 5, pp. 241–247, 2018.

[20] M. P. Starr, J. M. Schmidt (auth.), H. Stolp, et al., *The prokaryotes a handbook on habitats, isolation and identification of bacteria*, Springer-Verlag Berlin Heidelberg New York, 1st edition, vol. 1, 1981.

[21] M. Dworkin, S. Falkow, E. Rosenberg, et al., *The prokaryotes*, Springer New York, New York, NY, 3rd edition, vol. 7, 2006.

[22] E. van Eijk, B. Wittekoek, E. J. Kuijper, et al., “DNA replication proteins as potential targets for antimicrobials in drug-resistant bacterial pathogens,” *The Journal of antimicrobial chemotherapy*, vol. 72, no. 5, pp. 1275–1284, 2017.

[23] J. Bohlin, V. Eldholm, J. H. O. Pettersson, et al., “The nucleotide composition of microbial genomes indicates differential patterns of selection on core and accessory genomes,” *BMC genomics*, vol. 18, no. 1, p. 151, 2017.

[24] T. Hoffmann, D. Krug, N. Bozkurt, et al., “Correlating chemical diversity with taxonomic distance for discovery of natural products in myxobacteria,” *Nature communications*, vol. 9, no. 1, pp. 803, 2018.

[25] C. Vieille and G. J. Zeikus, “Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability,” *Microbiology and molecular biology reviews: MMBR*, vol. 65, no. 1, pp. 1–43, 2001.

[26] S. Mann and Y.-P. P. Chen, “Bacterial genomic G+C composition-eliciting environmental adaptation,” *Genomics*, vol. 95, no. 1, pp. 7–15, 2010.

[27] S. Thiery and C. Kaimer, “The predation strategy of *Myxococcus xanthus*,” *Frontiers in microbiology*, vol. 11, article 2, 2020.

[28] K. I. Mohr, “Diversity of myxobacteria-we only see the tip of the Iceberg,” *Microorganisms*, vol. 6, no. 3, pp. 1-23, 2018.

[29] B. Silakowski, B. Kunze, and R. Müller, “Multiple hybrid polyketide synthase/non-ribosomal peptide synthetase gene clusters in the myxobacterium Stigmatella aurantiaca,” *GENE an international journal on genes and genomes*, vol. 275, pp. 233–240, 2001.

[30] H. B. Bode, B. Bethe, R. Höfs, et al., “Big effects from small changes: possible ways to explore nature's chemical diversity,” *Chembiochem: a European journal of chemical biology*, vol. 3, pp. 619–627, 2002.

[31] N. Gaitatzis, B. Kunze, and R. Müller, “Novel insights into siderophore formation in myxobacteria,” *Chembiochem: a European journal of chemical biology*, vol. 6, no. 2, pp. 365–374, 2005.

[32] K. Gerth and R. Müller, “Moderately thermophilic myxobacteria: novel potential for the production of natural products isolation and characterization,” *Environmental microbiology*, vol. 7, no. 6, pp. 874–880, 2005.

[33] K. Gerth, S. Pradella, O. Perlova, et al., “Myxobacteria: proficient producers of novel natural products with various biological activities--past and future biotechnological aspects with the focus on the genus *Sorangium*,” *Journal of biotechnology*, vol. 106, 2–3, pp. 233–253, 2003.

[34] K. J. Weissman and R. Müller, “A brief tour of myxobacterial secondary metabolism,” *Bioorganic & medicinal chemistry*, vol. 17, no. 6, pp. 2121–2136, 2009.

[35] H. Hyun, S. Lee, J. S. Lee, et al., “Genetic and functional analyses of the DKxanthene biosynthetic gene cluster from *Myxococcus stipitatus* DSM 14675,” *Journal of microbiology and biotechnology*, vol. 28, no. 7, pp. 1068–1077, 2018.

[36] D. Krug, G. Zurek, B. Schneider, et al., “Efficient mining of myxobacterial metabolite profiles enabled by liquid chromatography-electrospray ionisation-time-of-flight mass spectrometry and compound-based principal component analysis,” *Analytica chimica acta*, vol. 624, no. 1, pp. 97–106, 2008.

[37] R. O. Garcia, D. Krug, and R. Müller, “Chapter 3: discovering natural products from myxobacteria with emphasis on rare producer strains in combination with improved analytical methods,” *Methods in enzymology*, vol. 458, pp. 59–91, 2009.

[38] J. Korp, L. Winand, A. Sester, et al., “Engineering pseudochelin production in *Myxococcus xanthus*,” *Applied and environmental microbiology, American society for microbiology*, vol. 84, no. 22, pp. 1–12, 2018.

[39] B. Kunze, F. Sasse, H. Wieczorek, et al., “Cruentaren A, a highly cytotoxic benzolactone from myxobacteria is a novel selective inhibitor of mitochondrial F1-ATPases,” *FEBS letters*, vol. 581, no. 18, pp. 3523–3527, 2007.

[40] D. R. Boone, R. W. Castenholz, and G. M. Garrity, *The Proteobacteria*, in Bergey's manual of systematic bacteriology, Springer, New York, 2nd edition, vol. 2, part C, 2005.

[41] A. D. Morgan, R. C. MacLean, K. L. Hillesland, et al., “Comparative analysis of *Myxococcus* predation on soil bacteria,” *Applied and environmental microbiology*, vol. 76, no. 20, pp. 6920–6927, 2010.

[42] R. O. Garcia, M. Stadler, and R. Müller, *Exploiting the natural products of novel myxobacteria: phylogenetic and fatty acid perspectives and bioactive compound discovery*: *Sandaracinus amylolyticus gen. nov., sp. nov., a starch-degrading soil myxobacterium, and the description of Aetherobacter gen. nov.*, Dissertation, Universität des Saarlandes, Germany, 2011.

[43] S. Meliah and P. Lisdiyanti, “Isolation, characterization and molecular identification of myxobacteria from two outermost islands of Indonesia,” *BIOTROPIA*, vol. 25, no. 2, pp. 121–129, 2018.

[44] S. Meliah, D. I. Kusumawati, and M. Ilyas, “Preliminary study of myxobacteria as biocontrol agents for panama disease pathogen, tropical race 4 *Fusarium odoratissimum*,” *IOP conference series: earth and environmental science*, vol. 457, pp. 1–8, 2020.

[45] E. Rosenberg, E. F. DeLong, S. Lory, et al., *The prokaryotes*: *deltaproteobacteria and epsilonpoteobacteria*, Springer Berlin Heidelberg, Berlin, Heidelberg, 4th edition, 2014.

[46] W. Dawid, “Biology and global distribution of myxobacteria in soils,” *FEMS microbiology reviews*, vol. 24, pp. 403–427, 2000.

[47] S. Schneiker, O. Perlova, O. Kaiser, et al., “Complete genome sequence of the myxobacterium *Sorangium cellulosum*,” *Nature biotechnology*, vol. 25, no. 11, pp. 1281–1289, 2007.

[48] K. I. Mohr, C. Wolf, U. Nübel, et al., “A polyphasic approach leads to seven new species of the cellulose-decomposing genus *Sorangium, Sorangium ambruticinum* sp. nov., *Sorangium arenae* sp. nov., *Sorangium bulgaricum* sp. nov., *Sorangium dawidii* sp. nov., *Sorangium kenyense* sp. nov., *Sorangium orientale* sp. nov. and *Sorangium reichenbachii* sp. nov,” *International journal of systematic and evolutionary microbiology*, vol. 68, no. 11, pp. 3576–3586, 2018.

[49] J. Korp, M. S. Vela Gurovic, and M. Nett, “Antibiotics from predatory bacteria,” *Beilstein journal of organic chemistry*, vol. 12, pp. 594–607, 2016.

[50] L. Burbank, M. Mohammadi, and M. C. Roper, “Siderophore-mediated iron acquisition influences motility and is required for full virulence of the xylem-dwelling bacterial phytopathogen *Pantoea stewartii* subsp. *stewartii*,” *Applied and environmental microbiology*, vol. 81, no. 1, pp. 139–148, 2015.

[51] B. C. Chu, A. Garcia-Herrero, T. H. Johanson, et al., “Siderophore uptake in bacteria and the battle for iron with the host; a bird's eye view,” *Biometals: an international journal on the role of metal ions in biology, biochemistry, and medicine*, vol. 23, no. 4, pp. 601–611, 2010.

[52] P. Meiser, K. J. Weissman, H. B. Bode, et al., “DKxanthene biosynthesis-understanding the basis for diversity-oriented synthesis in myxobacterial secondary metabolism,” *Chemistry & Biology*, vol. 15, no. 8, pp. 771–781, 2008.

[53] S. C. Wenzel and R. Müller, “The impact of genomics on the exploitation of the myxobacterial secondary metabolome,” *Natural product reports*, vol. 26, no. 11, pp. 1385–1407, 2009.

[54] H. Irschik, M. Kopp, and R. Müller, *Genes encoding the synthetic pathway for the production of Disorazole*, World Intellectual Property Organization, WO 2006/075013 A1, 2006.

[55] R. D. Noar and M. E. Daub, “Bioinformatics prediction of polyketide synthase gene clusters from *Mycosphaerella fijiensis*,” *PloS one*, vol. 11, no. 7, pp. 1–31, 2016.

[56] B. Silakowski, G. Nordsiek, B. Kunze, et al., “Novel features in a combined polyketide synthase/non-ribosomal peptide synthetase: the myxalamid biosynthetic gene cluster of the myxobacterium *Stigmatella aurantiaca* Sga151,” *Chemistry & Biology*, vol. 8, pp. 59–69, 2001.

Chapter 3. Polyphasic taxonomy of strain 1932KM as a proposed novel gliding bacterium in the phylum *Bacteroidetes*

Strain 1932KM has been identified as a novel strain. We examined the phenotype and genotype of this strain in-depth as part of a polyphasic taxonomy analysis. Also, we conducted a gene cluster analysis and bioassay against pathogenic bacteria and statistical analysis of 16S rRNA gene sequences from this strain with the statistical tool Palaeontological Statistics (PAST). Bioinformatic tools, such as Prokka, RAST, AntiSMASH, NP.searcher, PubChem, Simple Modular Architecture Research Tool (SMART), and Molinspiration, were used to conduct in-depth analyses of the whole-genome sequence.

3.1 Background

The term ‘polyphasic taxonomy’ was introduced 50 years ago by Colwell (1970). Taxonomy combines many levels of information, from molecular to ecological. It incorporates several distinct and separable portions of information extractable from a nonhomogeneous system to yield a multidimensional taxonomy [1,2]. Taxonomy has the same meaning as biosystematics, a process to figure out the classification, nomenclature and identification of unknown species [3]. A complete definition of biosystematics requires two additional parts, namely phylogeny and population genetics [3]. These describe the natural relationship between different bacteria in a phylogenetic relationship, as encoded in 16S rRNA sequence data [3]. The16S rRNA gene sequences themselves are the first step in examining a microorganism’s characteristics from molecular biology's viewpoint as part of a modern taxonomic study.

The gene sequencing of 16S rRNA makes up part of a polyphasic taxonomic microbial identification approach by providing information on genotypic similarity [4,5]. Genotypic information offers a gold standard for polyphasic taxonomy. The information is derived from nucleic acids like DNA and RNA or whole-cell proteins of the isolated organism [6]. DNA information has become a routine procedure in bacterial taxonomy to define phylogenetic relationships [1].

Therefore, a systematic process to identify and recognise a new species can use a combination of 16S similarity and other analyses [7]. DNA-DNA hybridisation and average nucleotide identity–OGRI (Overall Genome Related Index) can be used together with 16S similarity. This two-step approach is required to compare the type strain if the similarity between the sequences is ≥ 98.7% [7].

A polyphasic taxonomic approach is advantageous because it simultaneously exploits conventional and molecular identification techniques [4]. The major advantage of the polyphasic approach is that it can take the shared homologies of an unknown strain into account in assessing certain groups of bacteria as case studies to arrive at a consensus approach to microbial identification [4]. This approach includes phenotypic, chemotaxonomic and genotypic data and phylogenetic information [8]. This is where modern taxonomy makes use of certain approaches from conventional taxonomy.

Conventional taxonomy makes use of morphology, physiology, biochemical feature and chemotaxonomy properties. Morphology includes shape, length, colony colour, size, Gram-staining, pigment production, and gliding motility [3–5]. The physiological and biochemical features include data on growth rates at different temperatures, pH values, salt concentrations, atmospheric conditions, growth in various substances such as antimicrobial agents, and data on various enzymes' presence or activity and their metabolisation compounds [3]. Meanwhile, chemotype implies chemotaxonomic properties of the organism, e.g. cell wall composition, exopolysaccharides, isoprenoid quinones, polyamines, mycolic acids, teichoic acids, pigments, polyamines, whole-cell protein, FA profile, and polar lipid profile [6,9]. Morphology, physiology, and biochemical features are categorised as phenotypic information [3].

Genotypic and phenotypic information can develop testable and ultimately phylogenetic hypotheses and provide the best estimates about relationships between organisms in evolutionary history [10]. Therefore, using a polyphasic approach involving a combination of molecular biology techniques and conventional microbiological methods is necessary to obtain a better understanding of microbial

diversity [4]. A polyphasic taxonomy study aims to prove a strain's novelty through phenotypic and genotypic analysis [1,3,6,7,9].

We have identified a new gliding bacterium by its phenotypic characteristics, especially gliding motility on agar surface during myxobacteria isolation. The strain we isolated is named 1932KM, and preliminary tests have shown that the strain is a member of the phylum *Bacteroidetes*. The phylum *Bacteroidetes* could be isolated with the same isolation methods as myxobacteria [11]. Phylum *Bacteroidetes* is characterised by distinct gliding motility, which occurs in various ecosystems, habitats, lifestyles, and physiologies [12].

The phylum *Bacteroidetes* (previously known as the *Cytophaga-Flavobacterium-Bacteroides* or CFB group) is one of the major phylogenetic lineages Bacteria [5]. A complete list of genera in the phylum *Bacteroidetes* according to LPSN (List of Prokaryotic names with Standing in Nomenclature) can be seen in Supplement 2 [13–16].

The nomenclature of the genus *Cytophaga* is family *Cytophagaceae*, order *Cytophagales*, class *Cytophagia*, phylum *Bacteroidetes*, and domain Bacteria. The order *Cytophagales* are all unicellular, gliding, Gram-negative bacteria. They have rod-shaped cells, short or long, delicate or stout, with tapering or rounded ends [17]. The typical colonies are spreading swarms, filmlike, producing brightly-coloured colonies in shades of yellow, orange, or brick red [17]. The organisms belonging to the *Cytophagales* may be aerobic, microaerophilic, capnophilic (CO2-requiring), or facultatively anaerobic organotrophs, many of them able to degrade biomacromolecules like proteins, chitin, pectin, agar, starch, or cellulose [17].

Despite the wealth of phenotypic information, an in-depth analysis into the genotype remains important in acquiring complete genetic information about a microbial strain. Bioinformatic tools support this kind of analysis. The strain homology is comparable to other strains to construct a phylogenetic tree. The process of phylogenetic tree reconstruction through NCBI BLAST and/or MEGA

begins with a BioEdit analysis. Other bioinformatics software can provide a deeper analysis of the results.

3.2 Hypothesis

The present study aims to establish the taxonomic position of the gliding bacterium strain 1932KM that was recovered from Bali Beach, Indonesia. Thus, we sought to determine its phylogenetic position using a polyphasic taxonomic approach that includes both phenotypic and genotypic analysis. These analyses characterised 1932KM, which we propose to represent a novel genus of the family *Flammeovirgaceae* within the phylum *Bacteroidetes*.

3.3 State of the art

Marine organisms and habitats–for example, sea squirts, sponges, shellfish, algae, sediments and seashore sand–are appropriate sources for the isolation of *Flammeovirgaceae* bacteria [18–22]. Yoon et al. designated the *Flammeovirgaceae* family in 2011 [20], primarily basing the phylogeny on 16S rRNA gene data. At that time, the family 'Flammeovirgaceae' within the phylum *Bacteroidetes* incorporated the genera *Flammeovirga*, *Flexithrix*, *Limibacter*, *Perexilibacter*, *Persicobacter*, *Rapidithrix,* and *Sediminitomix* [20]. In August 2017, the *Flammeovirgaceae* family contained the following proposed and/or confirmed genera: *Algivirga, Aureibacter, Cesiribacter, Fabibacter, Fabivirga, Flammeovirga, Flexithrix, Fulvitalea, Fulvivirga, Imperialibacter, Limibacter, Nafulsella, Marinoscillum, Marivirga, Perexilibacter, Persicobacter, Porifericola, Rapidithrix, Reichenbachiella, Roseivirga, Sediminitomix,* and *Tunicatimonas.*

Several authors proposed to replace the combination of DNA-DNA hybridisation and average nucleotide identity–OGRI with 16S similarity for the classification of a new species [7]. We propose a new step to identify a new species through conserved region analysis from 16S rRNA gene sequence homology and pairwise comparison with the closest strain according to phylogenetic tree reconstructions. Housekeeping genes will be used as a reference in performing multiple alignment analysis for that purpose. Statistical analysis will be used to obtain a robust

conclusion about the strain's novelty. The statistical nonparametric analysis will focus on the correlation between strain 1932KM and the closest type strains.

3.4 Materials and methods

3.4.1 Sample

Strain 1932KM is a gliding bacterium isolated during a myxobacteria isolation project in the laboratory of Microbial Strain Collection at the Helmholtz Centre for Infection Research, Germany, in 2017. This strain came from a rock face in Pandawa Beach, Bali–Indonesia. It was collected in 2016 at coordinates 8°50'43" S 115°11'7" E. Bali is a volcanic island located in the Indian Ocean and surrounded by coral reefs. The average year-round temperature stands around 30°C with a humidity level of about 85%. The isolation media and method used were identical to the ones described in myxobacteria isolation procedures.

3.4.2 Isolate purification and maintenance

The strain was isolated on water-agar with *Escherichia coli* bait [11], purified according to Reichenbach & Dworkin [11], and maintained in a CY/H-medium. Supplement 1 provides the medium composition. Yoon et al. used this bait-streaked agar technique to isolate *Orifericola rhodea*–another genus in the *Flammeovirgaceae* family [23]. The incubation temperature was 30°C. Liquid cultures were incubated on a rotary shaker at 160 rpm.

3.4.3 Polyphasic taxonomy of strain 1932KM

3.4.3.1 Phenotypic analysis of strain 1932KM

3.4.3.1.a Morphology

Swarming colonies on agar plates were studied and photographed using a stereomicroscope (Olympus SZX12), while vegetative cells were observed through phase-contrast microscopy (Zeiss AX10), photographed using an Axiocam MRC (Zeiss) camera, and further analysed with the AxioVision LE software. The gliding and swarming behaviours were studied in several agar plate media, namely VY/2 +

50% Artificial Sea Water (ASW; ATI, Germany), CY + 1.5% NaCl, and Marine broth (MB; Roth), P-, VY/2-, and CY-agar [11].

3.4.3.1.b Aerobe /anaerobe test

Anaerobic growth was tested using Anaerocult (Merck) on CY- and MB-agar plates at 30°C and 37°C for 14 days. Meanwhile, catalase activity was analysed by bubble production in a 3% (v/v) hydrogen peroxide solution [24–26]. Oxidase was also tested with test stripes (Bactident Oxidase, Merck, Germany). To analyse growth temperature, we incubated the strain on MB agar at various temperatures, specifically 4°C, 18°C, 20°C, 24°C, 30°C, 38°C, and 44°C. Observation for optimum pH observation was conducted at pH values of 4.5–9.0 at intervals of 0.5, incubated for 5 days at 30°C. Salt tolerance was tested on CY agar incubated at 30°C for five days, varying in salt concentration from 0 to 9.0% (w/v) NaCl in 0.5% intervals. Duplicates were tested, and the volume of the strain taken from the liquid culture was 20 µl.

3.4.3.1.c Enzymatic test

Enzymatic activities were analysed with API ZYM bioMétrieux kit. API ZYM is a semi-quantitative micro method that consists of a strip with 20 microwells (cupules) that contain the enzymatic substrates and its buffer. This base allows contact between the enzyme and the insoluble substrate. The strip contains synthetic substrates and is made of non-woven fibres. The enzymatic test medium was inoculated with a dense suspension of the organism and used to rehydrate the enzymatic substrate. The metabolic products that were produced during the incubation period were detected by adding reagents that changed colour in the event of a positive detection. The result could be read after 4–4.5 hours at 37°C (optimum temperature). However, when the samples were being compared, all the test parameters (time, temperature, growth media, and density of the suspension) must be the same. Also, the inoculated strip should not be placed in bright light. Supplement 3 presents the list of enzymes and substrates.

In addition to APY ZYM, the strain 1932KM was also analysed for enzymatic activity or the fermentation of carbohydrates using API Coryne. The enzymatic test media were inoculated with a solid suspension of organisms, which reconstituted the enzymatic substrate. Colour changes after the addition of reagents at the end of the incubation period indicated the presence of metabolic products. The fermentation tests were inoculated with an enriched medium (containing a pH indicator), which reconstituted the sugar substrates. Fermentation of carbohydrates resulted in acidification, which was detected by a spontaneous colour change in the pH indicator. The reactions were read according to the Reading Table, and the identification was obtained by referring to the Analytical Profile Index.

3.4.3.1.d Antibiotic resistance test

Antibiotic resistance against 13 antibiotics was tested on VY/2 agar at 30°C for five days as described at Mohr et al. [27]. The final concentrations were set to 100 $\mu g.mL^{-1}$ for ampicillin, 30 $\mu g.mL^{-1}$ for chloramphenicol, 10 $\mu g.mL^{-1}$ for oxytetracycline and 50 $\mu g.mL^{-1}$ for the remaining antibiotics. For the negative resistance blank, the test strain was grown on VY/2 agar without antibiotics and incubated for 7 days at 30°C.

3.4.3.1.e Chemotaxonomic analysis

e.1 Menaquinone analysis

The menaquinone analysis method was adapted from Minnikin et al. (1984) and Wink et al. (2017) [28,29]. This method has three steps of analysis, namely menaquinone extraction, HPLC and LC-MS analysis. For menaquinone extraction, we prepared dry biomass (100 mg) in a polytetrafluoroethene capped tube, adjusted with 4 mL methanol−0.3% aqueous NaCl (90:10) and 4 mL petroleum ether. Then the sample was mixed on a tube rotator for 30 min. After that, we transferred the upper phase to a 25 mL evaporator. We added 4 mL of petroleum ether to the remaining sample (lower phase) and mixed it in a tube rotator for 30 min. Then we transferred the upper phase to the previous 25 mL evaporator tube, which already contains a solution from the previous process. After that, the solution was

evaporated at 30°C in a rotavorator until only 1 mL remained out of the original volume. In the next step, the solution was transferred to a 4 mL brown vial tube and dried at 30°C under N_2 pressure. Then the extract was re-dissolved in 100 µL acetonitrile-isopropanol (65:35) and transferred to the brown HPLC vial. It was kept at −20°C before analysis.

We used analytical reverse-phase high performance liquid chromatography (RP-HPLC) with an Agilent 1260 HPLC system with a diode array UV detector in the HPLC analysis. HPLC conditions: XBridge C_{18} column 100x2.1 mm (Water), 3.5 µm. Solvent A: H_2O-acetonitrile (95/5), 5 mmol NH_4AC, 0.04 $mL.L^{-1}$ CH_3COOH. Solvent B: H_2O-acetonitrile (5/95), 5 mmol NH_4AC, 0.04 $mL.L^{-1}$ CH_3COOH. Gradient system: 10% B increasing to 100% B in 30 min; flow rate 0.3 $mL.min^{-1}$; 40°C; Injection volume 5 µL, UV detection at 210–450 nm.

For the HPLC-HRESIMS analysis, the HPLC system was solvent A: Isopropanol, solvent B: Acetonitrile. Volume Injection 5 µL. Column C_{18}. Flow rate 0.3 $mL.min^{-1}$. The gradient system was 35% A and 65% B in 20 min with a waiting time of 10 min. UV detection at 255 or 270 nm. Pressure 400 bar. HRESIMS data were recorded on a MaXis ESI-TOF-MS spectrometer (Brucker).

e.2 Fatty acid analysis

Strain 1932KM was cultivated in a 300 mL flask containing 50 mL MB medium (Roth) shaken at 180 rpm for 4 days at 30°C. The cultures were harvested by centrifugation and used for FA extraction using the fatty acid methyl ester (FAME) method [30]. GC-MS analysis and identification of FAs were performed according to Gemperlein et al. [31]. The amounts of FA were calculated in percentages using the value obtained from the integrated signal.

3.4.3.2 Genotypic analysis

The DNA is an essential substance of genotypic analysis. In this research, we extracted the DNA of strain 1932KM throughout the same procedure on genomic DNA extraction as described in Chapter 2. We used the protocols from the

Invisorb® Spin Plan Mini Kit (250) manual, Stratec Molecular, REF 1037100300 to perform DNA isolation.

DNA analysis was done in the GMAK sequencing centre of HZI, Germany. The Sanger sequencing method was used as a method for analysing the 16S rRNA gene sequence. The universal primers of the 16S rRNA gene sequence were F27, R518, F1100, R1100 and R1492/R1525 at 10 pmol of concentration. Meanwhile, for whole-genome sequencing, the DNA sequencing library was generated from 200 ng DNA using NEBNext Ultra II FS DNA Library Prep Kit for Illumina (New England BioLabs) according to manufacturer's protocols including PCR Amplification with 4 cycles. The libraries were sequenced on Illumina MiSeq using MiSeq Reagent Kit v3 (600 cycles) with an average of 2,5 $x10^6$ reads per DNA sample. The genome of strain 1932KM was assembled with a unicycle (de novo assembly) [32].

The bioinformatics tools we used for DNA analysis were BioEdit [33], MEGA [34], PAST [35] and NCBI BLAST. These tools were used for 16S rRNA gene sequence analysis. We also used Prokka [36] to annotate whole-genome sequences and RAST [37–39] for a whole-genome sequence feature analysis. Furthermore, CGView [40–42] was used for whole-genome sequence visualisation, Phylosuite [43] for gene reference concatenation, and EZBioCloud [44] for OrthoANIu analysis. Finally, we also used NP.search, SMART and AntiSMASH [45] for gene cluster analysis.

3.4.4 Bioassay of antiinfectives

Finding and isolating an active compound against pathogens is a desirable goal to solve the global issue of antibiotic resistance. The method for this section refers to the bioassay method described in Chapter 2.

Strain 1932KM was fermented in 10 media for 5 days at 30°C, before extraction by methanol and subsequent screening of antiinfectives against 9 pathogenic bacteria. When the serial dilution test result achieved more than a three-fold dilution, the process was continued into compound analysis using HPLC and HPLC-HRESIMS.

3.5 Results and Discussion

3.5.1 Genotypic analysis

Strain 1932KM was picked out from the Indonesian samples during myxobacteria isolation. Strain 1932KM is a gliding bacterium with yellow, non-motile, strictly aerobic, rod-shaped cells and it liquefies agar. Full sequence 16S rRNA gene analysis showed that the strain is closest to the family *Flammeovirgaceae*. The nomenclature of the family *Flammeovirgaceae* places it in the phylum *Bacteroidetes*, class *Cytophagia*, order *Cytophagales*. The full 16S gene sequence of strain 1932KM has a shared similarity homology of 91.3% with the type strain *Rapidithrix thailandica* (DSM 103551^{T}; LC 191863) and 90.9% with *Flexithrix dorotheae* (DSM 6795^{T}; NR 040919.1). If the 16S similarity is below 98.7%, the strain can be categorised as a new species. Phylogenomic treeing should be used to recognise new genera [7]. Figure 3.1 shows the phylogenetic tree reconstruction for strain 1932KM.

The phylogenetic tree of strain 1932KM in Figure 3.1 is categorised as an unrooted tree. An unrooted tree does not require information about ancestry. The relationship between the strains is shown by distance in the clade. Evolutionary analysis of the strain 1932KM against 50 type strains based on NCBI databases was inferred using the NJ (Neighbor-Joining) method [34]. The Neighbor-Joining method's principle is that neighbours are defined as two taxa connected by a single node in an unrooted tree [46]. If the two strains' branches' lengths are not too different, the strains should not be very different either. A longer branch signifies more genetic change. The branching of strain 1932KM was closely related to *Rapidithrix thailandica* (DSM 103551^{T}; LC 191863.1) and *Flexithrix dorotheae* (DSM 6795^{T}; NR 040919.1).

Genomic comparison between a strain and a type of strain is the way to obtain a robust conclusion about a novel species. Novelty can be assumed when the OGRI calculations result in an ANI (Average nucleotide identitiy) of < 95–96% or DDH (DNA-DNA hybridisation) of 70% [7]. In this case, strain 1932KM was compared

to *Flexithrix dorotheae* (DSM 6795^T; NR 040919.1). The Fasta file of DSM 6795^T with Taxonomy ID 1121904 was downloaded from www.ezbiocloud.net.

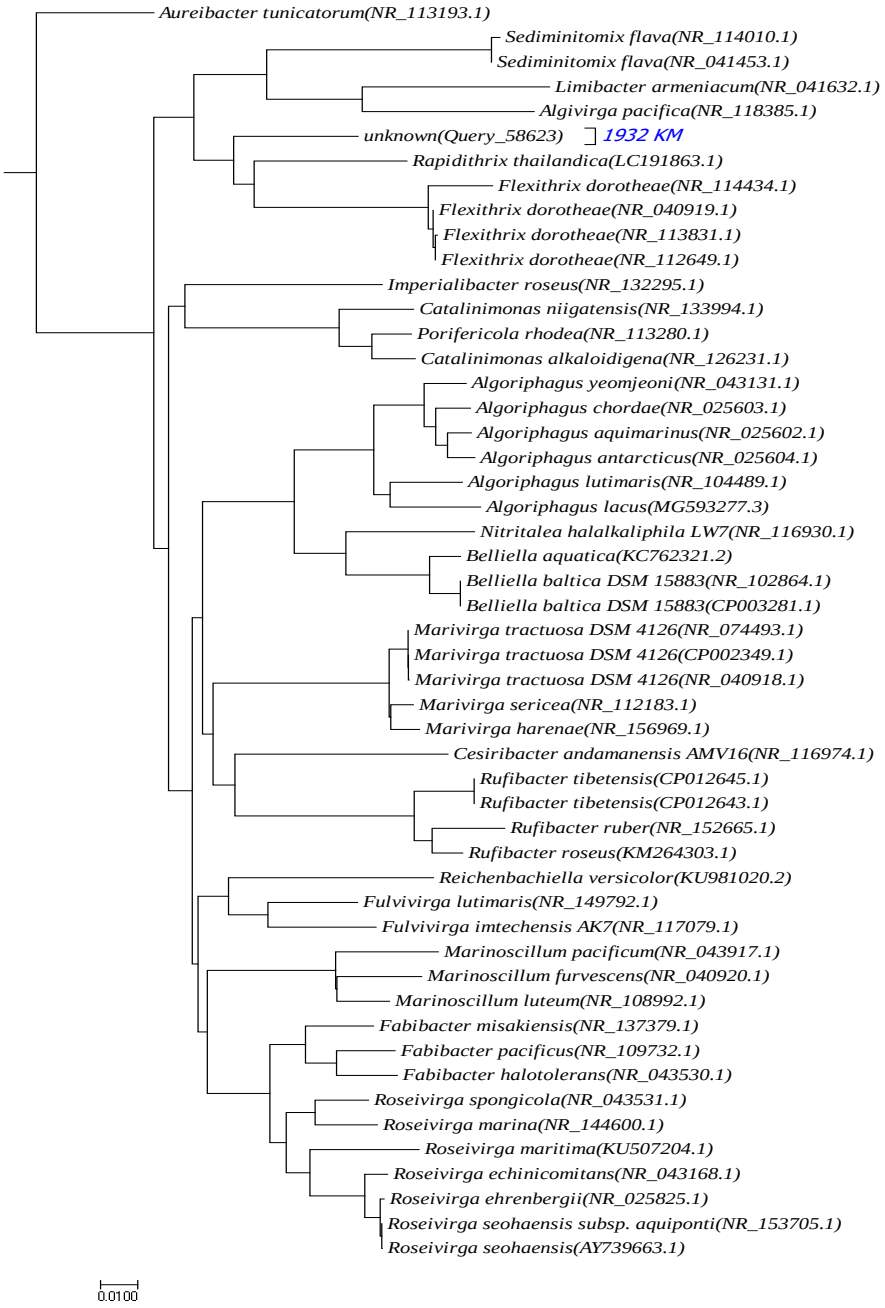


Figure 3.1 Phylogenetic tree reconstruction according to similarities to NCBI results by Mega7.

The average nucleotide identity (ANI) is defined as a pairwise measure of overall similarity between two genome sequences [44]. In the ANI calculation, the suggested species boundary is 95–96% [44]. Yoon et al. (2017) developed a new ANI calculation, namely the orthologous ANI algorithm using the USEARCH (OrthoANIu) programme. This tool was developed by Edgar (2010) as a new algorithm to enable sensitive local and global searches of large sequence databases at exceptionally high speeds [47]. OrthoANIu has been compared to other ANI algorithm programmes, resulting in the conclusion that OrthoANIu does not produce different values but has the fastest data processing [44].

OrthoANIu is suitable across the full range of ANI calculations while being computationally efficient enough for use in large-scale comparative genomic studies [44]. According to the analysis results from http://www.EZBioCloud.net/tools/ani, the OrthoANIu value of strain 1932KM against DSM 6795^T is 69.90%. This value shows that strains 1932KM and DSM 6795^T are not identical and that the former can be categorised as a novel strain.

As we have mentioned before, DDH is the taxonomic gold standard for species delineation in *Archaea* and *Bacteria* [48]. Genome Blast Distance Phylogeny produced a DDH estimation for 1932KM compared to DSM 6795^T of 19.60% with a distance of 0.2237. Distances are inferred using three distinct formulas from the set of HSPs MUMs obtained by comparing each genome pair with the chosen software [48]. These distances are transformed to values analogous to DDH using a generalised linear model inferred from an empirical reference dataset comprising real DDH values and genome sequences [48]. According to this result, strain 1932KM and DSM 6795^T are distinct species.

We also performed analysis with other bioinformatics tools. The aim is to analyse the differences and similarities between the genomic features of strains 1932KM and DSM 6795^T. Annotation of strain 1932KM has been done by Prokka [36]. Meanwhile, genomic information of DSM 6795^T was obtained from NCBI with ID 14229 and EZBioCloud.

Analysis into the whole-genome sequence of strain 1932KM by Prokka identified that this strain contains 87 contigs, 8,681,004 bases, 6.689 CDS, 3 rRNA, 2 repeat regions, 46 t-RNA and 1 t-mRNA. As well into the whole-genome sequence of type strain DSM 6795^{T} by Prokka identified that this strain contains 101 contigs, 9,536,821 bases, 7,304 CDS, 6 rRNA, 1 repeat region, 44 t-RNA and 1 t-mRNA. Deeper analysis requires other tools explained above. As for the terminology, contigs are a set of sequences that overlap and contain genetic information. A base pair is an essential unit in double-stranded nucleotides, which has hydrogen bonds connecting bases. CDS or coding sequence is the part of genes that is translated into proteins. Ribosomal ribonucleic acid (rRNA) is a ribonucleic acid molecule in ribosomes that is not translated into protein. Repeat region signifies multiple copies of a region in the genome. Transfer RNA is a molecule with an important role in the translation of genetic code into new proteins. A transfer-messenger RNA is an RNA molecule that has t-RNA and m-RNA properties.

The subsystem features of DSM 6795^{T} and strain 1932KM were assessed through RAST (Rapid Annotations using Subsystem Technology). The differences between the strains provide information on strain characteristics for taxonomic clustering. Genomic information on strains 1932KM and DSM 6795^{T} can be seen in Table 3.1.

Genomic profiles of the strains 1932KM and DSM 6795^{T} show that both strains exhibit similarities in subsystem feature counts for ‘phages, prophages, transposable elements, plasmids’, ‘dormancy and sporulation,’ and ‘cell division and cell cycle.’ In the description of ‘phages, prophages, transposable elements, plasmids’, part of the subsystem was found on a pathogenicity island. The system found that the query strain’s gene displays similarity to *Listeria’s* Pathogenicity Island LIPI-1 extended. Listeria's main pathogenicity island LIPI-1 carries virulence genes essential for intracellular parasitism [39,49,50]. Based on this similarity, strain 1932KM is capable of infecting other organisms. Meanwhile, for ‘dormancy and sporulation’, several proteins were identified due to their sporulation effects, although they also have broader cellular roles [39].

In the description for ‘cell division and cell cycle,’ the subsystem shows a similarity with the bacterial check-point-control-related cluster. It controls the protein DisA with di-adenylate cyclase activity, a hypothetical protein and phosphoglucosamine mutase (glmM) [39].

Table 3.1 Genomic information on strains 1932KM and DSM 6795^{T}

Organisms overviews and sub-system feature counts	Strain 1932KM	DSM 6795^{T}
No. of CDSs	6,681	7,391
No. of contigs	87	101
rRNA	3	6
tRNA	46	45
Size (Mbp)	8.68	9.54
GC content (%)	39.6	36.3
Cofactors, vitamins, prosthetic groups, pigments	194	191
Cell wall and capsule	36	44
Virulence, disease, and defense	43	57
Potassium metabolism	10	17
Photosynthesis	0	0
Miscellaneous	13	18
Phages, prophages. transposable elements, plasmids	1	1
Membrane transport	85	90
Iron acquisition and metabolism	12	10
RNA metabolism	47	51
Nucleosides and nucleotides	61	60
Protein metabolism	161	124
Cell division and cell cycle	3	3
Motility and chemotaxis	0	0
Regulation and cell signalling	30	47
Secondary metabolism	8	7
DNA metabolism	57	59
Fatty acids, lipids, and isoprenoids	22	48
Nitrogen metabolism	14	28
Dormancy and sporulation	1	1
Respiration	62	81
Stress response	26	48
Metabolism of aromatic compounds	14	27
Amino acids and derivatives	275	288
Sulfur metabolism	46	99
Phosphorus metabolism	21	32
Carbohydrates	265	310

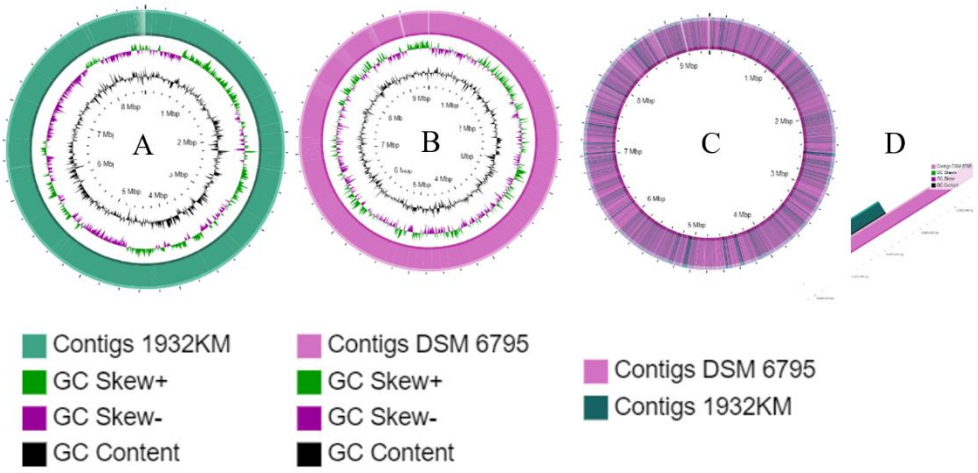
The octameric DisA protein complex has structurally linked nucleotide-binding. DNA-binding HhH domains and the nucleotide-binding domains are bound to cyclic di-adenosine phosphate such that DisA is a specific di-adenylate cyclase [39]. The di-adenylate cyclase activity is strongly suppressed by binding to branched DNA, but not to duplex or single-stranded DNA, suggesting a role for DisA as a monitor of the presence of stalled replication forks or recombination intermediates via DNA structure-modulated c-di-AMP synthesis [39]. The resulting cyclic diadenosine phosphate, c-di-AMP, is reminiscent but distinct from c-di-

GMP, a prokaryotic regulator of complex cellular processes [39]. The size of the 'uncharacterised secreted protein associated with spyDAC' is very variable [39].

Phosphoglucosamine mutase (glmM) catalyses the interconversion of glucosamine-6-phosphate into glucosamine-1-phosphate, an essential step in the biosynthetic pathway leading to the formation of the peptidoglycan precursor uridine 5'-diphospho-N-acetylglucosamine [39]. An insertional glmM mutant of *S. gordonii* did not produce GlmM and had a growth rate that was approximately half that of the wild type [39]. Morphological analyses indicated that the glmM mutation causes marked elongation of the streptococcal chains, enlargement of bacterial cells and increased roughness of the bacterial cell surface. Furthermore, the glmM mutation reduces biofilm formation and increases sensitivity to penicillin [39]. Compared to the wild type, glmM mutation resulted in increased sensitivity to PMN-dependent killing [39]. GlmM may be required to synthesise firm peptidoglycans for resistance to bacterial cell lysis [39]. Vice versa, the glmM mutation may be involved in breaking peptidoglycans in the process of bacterial cell lysis. Blocking one of the biosynthetic pathways in the peptidoglycan formation is a potential target against bacterial resistance.

Whole-genome comparisons are an important method of analysis to obtain a better understanding of organisms. Figure 3.2 is a visualisation of the genomic comparisons between strains 1932KM and DSM 6795^{T} according to the web-service analysis at http://cgview.ca.

Figure 3.2 Visualisation of BLAST genomic comparison between 1932KM and DSM 6795^{T}



Visualisation of genomic comparison allows information from various sources to be integrated and interpreted [41]. The numbers of contigs (sequences) for strain 1932KM and DSM 6795^{T} are 8.68 Mbp and 9.54 Mbp, respectively, as shown in the outermost circle. The explanation for these displays above is adapted from the genome map in EZBioCloud. When all the BLAST sequences are shown next to each other (Figs. C and D), visualisation makes it easier to understand.

According to the EZBioCloud explanation, the GC skew matrix in (A) and (B) is used as an indicator for identifying replication loci and leading/lagging strands. The genomic mean GC-skew value is used as the baseline. Higher-than-average values (GC Skew+) are displayed in **green**, and lower-than-average (GC Skew-) values are displayed in **purple**.

The displays of the GC ratio matrix can be used to profile the genome, identify isochores, or observe co-variations with other data. An isochore is a large region of DNA (> 300 kb) with a high degree of uniformity in guanine (G) and cytosine (C): G-C and C-G), which is collectively known as G+C content [51,52]. The GC ratio uses the genomic mean GC ratio value as its baseline, with higher-than-average values and lower-than-average values shown in black. The G+C contents of the strains 1932KM and DSM 6795^{T} are 39.6% and 36.3%, respectively.

3.5.1.1 Application of housekeeping genes in 16S rRNA gene sequence analysis of conserved regions, identity matrix scores, and single nucleotides

Sequence comparison is intended to find differences and similarities between strains. All the methods present their results as a number, percentage, or visualisation. These methods require whole-genome sequences to analyse as well. There are no methods that examine differences between strains down to the position of individual nucleotides in the residue. 16S rRNA gene sequences analysis, in the beginning, shows a percentage of similarity-based pairwise alignment between the query and type strains.

In this dissertation, we would like to conduct an in-depth analysis of the full 16S rRNA gene sequences of strain 1932KM. The goal is to determine the number of conserved regions and the identity matrix score for strain 1932KM and compare these results to the two closest type strains. The positions of single nucleotides could be detected as well. Still, due to the percentage of similarity we have mentioned previously (less than 98.7%), the single nucleotides would be widely spread and make up a long list. Finding characteristics among them is necessary to obtain a robust and final conclusion on the nomenclature of strain 1932KM. There is no standardised or consensus guideline to establish the case definition for the classification of a species or genus using partial 16S rRNA gene sequencing [53].

Conserved regions and identity matrix scores can result in multiple alignments between three sequences. This alignment involves a reference strain, strain 1932KM as the query strain and a type strain. As a reference, this analysis used three housekeeping genes, namely pgm, pyrG, and rpoB. Phosphoglucomutase (pgm), cytidine triphosphate synthetase (pyrG), and RNA polymerase β-subunit (rpoB) are genetic markers for organisms and presumably exist in many organisms. The advantage in using housekeeping genes as a reference is that it presents a light processing burden for a graphic operating system running on an ordinary computer (such as Windows, Linux, or OS), especially compared to using whole-genome sequences as a reference.

The concatenation of three housekeeping genes as a reference resulted in a sequence longer than the queried sequence. Multiple alignments should be examined first, along with the concatenation of reference genes. BioEdit contains tools to analyse conserved regions and identity matrix scores, but the concatenation required the use of Phylosuite software. Conserved regions and identity matrix scores were used to identify homologous sites within each gene. As a note, the parameters were set up according to the software templates.

The multiple alignment analysis between housekeeping genes as a reference for the strains 1932KM and DSM 6795^{T} in BioEdit produced a score of 0.333 for the identity matrix, 7 conserved regions and 122 single nucleotides. Previously, it has been mentioned that strain 1932KM is closest to the type strain *Rapidithrix thailandica* (DSM 103551^{T}; LC 191863.1). Unfortunately, genomic information about this strain could not be found, but the Fasta file of 16S rRNA gene sequences for the strain DSM 103551^{T} could be analysed with this method. Using the same workflow previously used with strains 1932KM and DSM 103551^{T}, the multiple alignment analysis resulted in a score of 0.303 for the identity matrix, 4 conserved regions and 109 single nucleotides. Table 3.2 presents a summary of this data.

Table 3.2. Resume of pairwise and multiple alignment analysis

Strains	**Length (bp)***	**Similarity (%)**	**Identity matrix**	**Conserved regions**	**Single nucleotides (pairs)**
1932KM	1488	-	-	-	-
DSM 103551^{T}	1446	91.27	0.303	4	109
DSM 6795^{T}	1478	90.92	0.333	7	122

*: bp = base pairs

The data shown above displays a correlation between the percentage of similarity between strain 1932KM and the type strains in 16S rRNA gene sequences. The variable analysis through pairwise comparison and multiple alignments on the one hand. When the percentage of similarity is at its lowest, the identity matrix displays

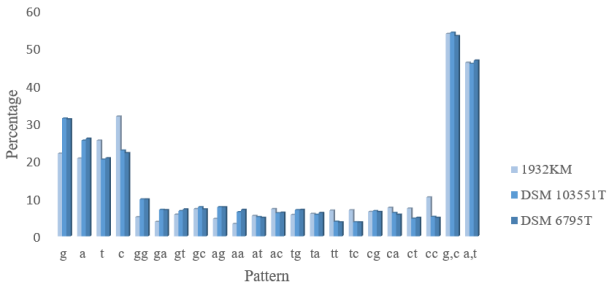
the lowest score as well. Meanwhile, the numbers of conserved regions and single nucleotide pairs are highest for the type strain that exhibits the highest similarity. All this data supports the notion of strain 1932KM as a novel strain with distinct dissimilarities to the type strains. In addition to the numbers of conserved regions, the identity matrix scores and single nucleotide analysis, the full 16S rRNA gene sequence analysis included statistical analysis as well.

3.5.1.2 Statistical analysis approach in 16S rRNA gene sequence analysis for strain identification

A statistical approach was included in this 16S rRNA gene sequence analysis for a robust analysis for strain novelty. Shared similarities between strain 1932KM and the type strains were found in a wide range of percentages. It was thus assumed that strain 1932KM had dissimilarities with the closest type strains. A statistical analysis approach should be applied to prove this hypothesis.

Firstly, the residue in sequences was converted using a statistics tool, the DNA Statistics at https://www.bioinformatics.org/sms2/dna_stats.html. DNA Statistics returns the number of occurrences of each residue in the sequence and percentage totals are given for each residue and certain groups of residues, which allowed us to quickly compare the results obtained for different sequences [54]. The data can be seen in Chart 3.1. The chart shows that the residues of strains have a narrow range of distances.

Chart 3.1. DNA statistics of 1932KM, DSM 103551^{T}, and DSM 6795^{T}



DNA Statistics tool converted the DNA residue into nonparametric data in percentage form. This represents residues singly or in pairs without any required variable parameters. Nonparametric statistics encompasses statistical methods that do not make any assumptions about the underlying distribution and independence of data [55].

The data distribution analysis proceeds into the hypothesis test using the percentages given for each strain. The data distribution in DNA statistics was analysed using the Epps-Singleton method for two-sample tests with the statistical tool PAST version 3.25. The result of the distribution test using Epps-Singleton for these strains is shown below in Table 3.3.

The Epps-Singleton test is a nonparametric test for the overall equal distribution of two univariate (one variable) samples. It is typically more powerful than the Kolmogorov-Smirnov test [35,56,57]. The statistic test W_2 is based on the difference between the two samples, and the system will apply the small-sample correction to W_2 if both sample sizes are less than 25 [35]. W_2 is a measure for the statistical distance between the empirical characteristics functions of both samples standardised by the variance-covariance underlying the two samples in question [57].

Table 3.3. Distribution test of 1932KM, DSM 103551^T and DSM 6795^T using Epps-Singleton

Sequence 1	Sequence 2	Equal test distribution, Epps-singleton, two samples test		
		N	W_2	p (same distribution)
1932KM	DSM 103551^T	22	0.0019	1
1932KM	DSM6795^T	22	0.0099	0.9999

Where: N = data size; W_2 = statistics testing; p= confidence of significance

From the results in Table 3.3, the pairwise comparison between strains 1932KM and DSM 103551^T produced a value of 0.0019 with a p-value of 1. Meanwhile, a comparison between strains 1932KM and DSM 6795^T gave several 0.0009 with a p-value of 0.9999. PAST interpreted the data for DNA statistics of strains 1932KM with DSM 103551^T and DSM 6795^T as indicating that the distribution is not

significantly different. The significance of the p-value of these same distributions was close to 1 (one). If we only had the data distribution result, it would not have been enough to determine the strains' correlation.

The correlation between the query and subject strains is shown in a correlation coefficient–a scale-invariant asymmetric measure of association between two random variables. It has a range from (-)1 to (+)1, where the extremes indicate perfect correlation and 0 (zero) means no correlation [58]. The sign is negative when a large value for one variable is associated with a small value for the other and positive if both variables tend to be large or small simultaneously [58]. The rule for interpreting the size of Correlation Spearman's coefficient is 0.90 to 1.00 (−0.90 to −1.00) indicates very high positive (negative) correlation; 0.70 to 0.90 (−0.70 to −0.90) high positive (negative) correlation; 0.50 to 0.70 (−0.50 to −0.70), moderate positive (negative) correlation; 0.30 to 0.50 (−0.30 to −0.50), low positive (negative) correlation) and 0.00 to 0.30 (−0.00 to −0.30), negligible correlation [59].

The Spearman correlation analysed the correlation of strain 1932KM to type strains DSM 103551^T and DSM 6795^T. Table 3.4 presents the result.

Table 3.4. Correlation testing between strain 1932KM with DSM 103551^T and DSM 6795^T using Spearman tests

Sequence 1	**Sequence 2**	**N**	**DOF**	**rs**	**D**	**D $_{table}$ $\alpha = 0.05$**
1932KM	DSM 103551^T	22	20	0.4224	0.1784	0.447
1932KM	DSM 6795^T	22	20	0.3992	0.1594	

Where, DOF: Degree of Freedom, rs: Correlation Spearman`s Coefficient, D: Determination Coefficient

From these results, the Spearman's coefficient for the correlation of strain 1932KM to DSM 103551^T and DSM 6795^T indicates low positive values. This conclusion refers to Spearman's Correlation coefficient. A low positive correlation is in the range of 0.30 to 0.50 [59]. The low positive correlation means that strain 1932KM appears to be quite different compared to both type strains.

Meanwhile, the coefficient of determination (D) indicates that the level of variance in the dependent variable (strain sequence) caused by its correlation to the

independent variable (pattern of DNA statistics) is higher in the comparison between strains 1932KM and DSM 103551^{T} than in that between strains 1932KM and DSM 6795^{T}. This is consistent with previous results that the homology between strains 1932KM and DSM 103551^{T} (91.3%) is higher than that between strains 1932KM and DSM 6795^{T} (90.9%).

The last part of Spearman's Correlation analysis is the significance test. This test is part of the analysis needed to approve or reject a hypothesis. In this analysis, we use a confidence interval of 95%, an α-value of 0.05 and a Degree of Freedom (DOF) of 20. In this case, the hypothesis H-null is accepted if strain 1932KM is dissimilar to the type strains with a Spearman's correlation (rs) < D table. Conversely, H-null is rejected if strain 1932KM is similar to the type strains with rs > D table. The result of the significance test is that the rs value for strains 1932KM and DSM 103551^{T} was less than the D table (rs < D table), which means that H-null is accepted. This result is the same for strains 1932KM and DSM 6795^{T}. The conclusion is that strain 1932KM exhibits similarities to both type strains, but it is significantly different.

Statistical analysis could be a part of the genotypic analysis, as demonstrated in the homology analysis of 16S rRNA gene sequences to whole-genome sequences. The next step in this analysis is to confirm our previous results. This should be done after BLAST for 16S rRNA gene sequencing analysis. This step is used to support a decision on the choice of a type strain as a basis for comparison to the unknown strain in order to achieve a robust conclusion.

A scale-invariant asymmetric was used as an observational parameter to see the extent of dissimilarity between the strains. Also, the correlation size produced by this method could be a consideration in categorising the novelty level of an unknown strain, e.g. a new species, genus, or family, etc.

Genotypic analysis works together with phenotypic analysis and chemotaxonomy under polyphasic taxonomy. The phenotypic observation included as part of the analysis involves morphology and physiology. Finding out the target strain's

morphology from the beginning would make it easier to recognise the strain during isolation processing.

3.5.2 Phenotypic analysis

Strain 1932KM was maintained in a CY/H-medium. The incubation temperature was 30°C. The liquid cultures were incubated in a rotary shaker at 160 rpm. Swarming colonies on agar plates were studied and photographed using a stereomicroscope (Olympus SZX12). At the same time, vegetative cells were observed by phase-contrast microscopy (Zeiss AX10), photographed with an Axiocam MRC (Zeiss) camera, and further analysed using the AxioVision LE software.

Strain 1932KM shows the typical gliding and swarming activity on solid agar surfaces and swarms extensively on agar plates (16 $g.L^{-1}$) of VY/2 + 50% Artificial Sea Water (ASW; ATI, Germany), CY + 1.5% NaCl [11], and Marine broth (MB; Roth; Figure 2a). The strain also grows well but swarms less on salt-free nutritious P- [60], VY/2-, and CY-agar [11]. On water-agar having North Sea water instead of deionised tap water, supplemented with *E. coli* bait-cross on the top, the strain grows over the bait without degrading it. On ASW-agar, thin, transparent swarms and few cell mass production can be observed. On VY/2 agar, the strain grows ribbed (Figure 2d). The colony is golden yellow on MB, VY/2 and CY + 1.5% NaCl-agar (Figure 2a, b and c), as well as in liquid CY and CY/H medium (Figure 2e) and transparent on H_2O- (North Sea water) + *E. coli*, VY/2 + 50 % ASW- and on ASW-agar. Weak starch degradation was visible on P agar, and no yeast degradation (clearing zone) could be observed on VY/2 agar. Agar liquefaction occurred on CY agar with 1.5 % NaCl (Figure 2b). Cells are rod-shaped with tapering ends (Figure 2f), 0.7 µm wide and 5.6 µm long.

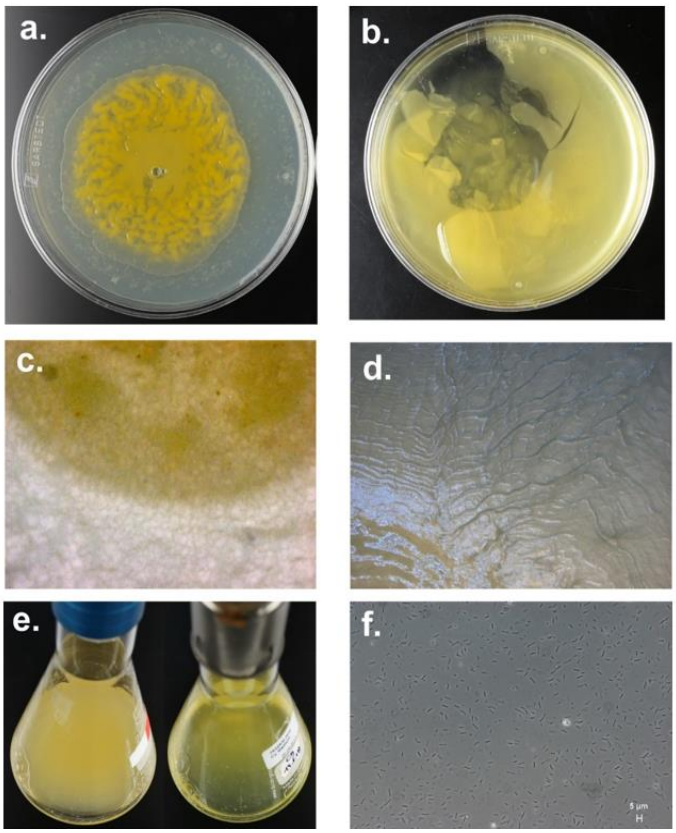


Figure 3.3 Profile of strain 1932KM in some conditions. Strain 1932KM on **a.** Marine Broth (MB), two weeks; **b.** CY agar liquefaction, four weeks; **c.** CY agar, nine days; **d.** Ribbed growth on VY/2 agar, two weeks; **e.** liquid medium CY/H (left) and CY (right), one week; **f.** Cells from liquid CY/H.

All cultures were incubated at 30°C.

However, the cell shape of the most closely related type strain *Rapidithrix thailandica* is 0.7 in width and 20–100 µm in length [61]. The cell shape of the genus *Flexithrix* (the second most closely related genus) is 0.4–0.9 µm wide and 1.5–70 µm long [11]. Growth in liquid cultures was homogenous. Anaerobic growth, tested in an anaerobic chamber using Anaerocult (Merck) on CY- and MB-agar plates at 30°C and 37°C for 14 days, respectively, was impossible. Microaerophilic growth was tested in CY prick-agar-tubes for a month at 30°C. Growth was possible on the surface of the puncture-channel. The cultured members of *Flammeovirgaceae* grow aerobically except for species of *Persicobacter*, which are facultative anaerobic [62,63].

Catalase activity, assessed by bubble production in a 3% (v/v) hydrogen peroxide solution [64], was negative, as described for *Rapidithrix* [61]. Oxidase, tested with test stripes (Bactident Oxidase, Merck, Germany), was positive, as described for the closely related genera *Rapidithrix* and *Flexithrix*.

Strain 1932KM grows at mesophilic temperatures and a high pH range. Growth at different temperatures was tested with 20 µl of liquid culture inoculated in the middle of marine broth agar. From plates incubated at 4°C, 18°C, 20°C, 24°C, 30°C, 38°C and 44°C, respectively, the swarm diameter was measured after five days. Growth was possible between 18°C and 38°C, with the best growth rate for strain 1932KM found at 38°C. No growth on agar plates occurred at 4°C and 44°C. At 18°C and 20°C, growth was weak; at 24°C and 30°C, good growth was possible. Growth at different pH values (4.5–9.0) in intervals of 0.5 was tested on marine broth agar inoculated with 20 µl of liquid culture. Based on the swarm diameter after five days at 30°C, the new strain 1932KM tolerated pH 5.0–9.0 and grew optimally at pH 5.0–8.0. However, no growth is possible at pH 4.5.

Salt tolerance was tested on CY agar at 30°C for five days with different salt concentrations from 0 to 9.0% (w/v) NaCl in intervals of 0.5%. Strain 1932KM tolerates NaCl-concentrations up to 6.0% and grows best at concentrations between 0 and 1.5%. Good growth was possible on concentrations between 2.0–5.0% and weak growth on agar with 5.5 and 6.0% NaCl. Duplicates were tested. Table 3.5. shows comparative characteristics of strain 1932KM and the most closely related genera and species *Rapidithrix thailandica* and *Flexithrix dorotheae*.

Table 3.5 Comparative phenotypic analysis between strain 1932KM and the type strains

Characteristics	**1932KM**	***Rapidithrix thailandica****	***Flexithrix dorotheae****
Cell morphology (wide/long; µm)	Rods (0.7/ 5.6)	Non-sheeted filaments (0.7/ 20-100)	Sheeted filaments (0.4-0.9/ 1.5-70)
Color of cell mass	Yellow	Light olive grey	Yellow
Catalase	-	-	+
Agar liquefaction	+	n.d.	-
Growth temperature (°C); optimum	18 - 38; 38	n.d.; 25 - 30	17 - 40; 35 - 40
pH tolerance; optimum	5.0 - 9.0; 7.5 - 8.0	5.0 - 10.0; n.d.	6.5 - 8.0; 7.0
Salt tolerance (%); optimum	0 - 6.0; 0 - 1.5	n.d.	2.0 -5.0 n.d.

*Data from Srisukchayakul et al. [61]: only optimum values are mentioned for temperature and pH; **Data from Reichenbach [65] and Hosoya and Yokota [66]

The API ZYM test showed strong activity for alkaline phosphatase, leucine arylamidase, valine arylamidase, acid phosphatase, α-galactosidase, α-glucosidase, and N-acetyl-β-glucosaminidase. However, esterase (C4), cystine arylamidase, trypsin, α-chymotrypsin, naphthol-AS-Bl-phosphohydrolase, β-galactosidase, and α-fucosidase activities were weak. Furthermore, esterase lipase (C8), lipase (C14), β-glucuronidase, β-glucosidase, N-acetyl-β-glucosaminidase, and α-mannosidase activities could not be identified. In the API Coryne test, strain 1932KM showed pyrazinamidase, alkaline phosphatase, β-galactosidase, α-glucosidase, and β-glucosidase activity, but no activity in nitrate reduction, pyrrolidonyl arylamidase, β-glucuronidase, N-acetyl-β-glucosaminidase, urease, or gelatine hydrolysis. Neither was there any fermentation of glucose, ribose, xylose, mannitol, maltose, lactose, saccharose, or glycogen. Table 3.6 summarises the test results.

Table 3.6 Profile of API ZYM and API Coryne results for 1932KM

Enzymes	API ZYM	API Coryne
Alkaline phosphatase	+	+
Leucine arylamidase	+	
Valine arylamidase	+	
Acid phosphatase	+	
Naphthol-AS-Bl-phosphohydrolase	+	
Esterase (C4)	+	
Esterase lipase (C8)	+	
Lipase (C14)	+	
Cystine arylamidase	+	
Trypsin	+	
Chymotrypsin	+	
α-galactosidase	+	
β-galactosidase	+	+
β-glucuronidase	n.d	n.d
α-glucosidase	+	+
β-glucosidase	n.d	+
N-acetyl-β-glucosaminidase	+	n.d
α-mannosidase	n.d	
α-fucosidase	+	
Reduction of nitrase		n.d
Pyrazinamide		+
Urease		n.d
Pyrrolidonyl arylamidase		n.d
Hydrolysis gelatin		n.d

API ZYM and API Coryne are general biochemical (i.e. enzymological) evaluation methods for assessing microbial functional diversity [67]. The reactions were not affected by atmospheric conditions or the type of media used in culturing the test organisms. Still, the colour intensity of the reactions varied to some extent with the bacterial test suspension [68]. API Coryne is a commercial system for identifying aerobic or facultative, non-branched and non-spore-forming Gram-positive bacteria [69]. These methods were easy to set up, requiring no special apparatus or reagents other than those supplied (because of the finding that ultrasonication resulted in the detection of fewer enzymes). Results for a wide variety of organisms were easy to read and repeatable [70]. The API system has successfully been used to produce an enzymatic profile useful in identifying many animal pathogens. It has also been used in the taxonomic classification and virulence evaluation of organisms [68]. Although API Coryne is intended for the identification of Gram-positive bacteria,

strain 1932KM shows a different profile. Enzyme β-glucosidase was detected by API Coryne but not by API ZYM. Conversely, N-acetyl-β-glucosaminidase activity was detected by API ZYM but not by API Coryne.

β-glucosidases, a heterogeneous group of exo-type glycosyl hydrolases, cleave β-glucosidic linkages in disaccharide or glucose-substituted molecules [71]. β-glucosidases occur in all kingdoms of life and play fundamental biological roles in processes such as the degradation of cellulose and other carbohydrates for nutrient uptake and developmental regulation or chemical defence against pathogen attack [71]. β-glucosidases from various biological sources open important avenues of exploration for further practical applications [71].

N-acetyl-β-glucosaminidase was detected through the API ZYM test and also by RAST as part of Subsystem 'chitin and N-acetylglucosamine utilisation in amino sugar and nucleotide sugar metabolism.' According to RAST results, glycolysis has an enzymatic point metabolism, with chitobiose as the input and N-acetyl-D-glucosamine 6-phosphate as the result. The metabolism continues to produce D-fructose 6-phosphate.

N-acetyl-β-d-glucosaminidase is one of the three enzymes that degrade chitin by catalysing the hydrolysis of chitin (N-acetyl-ß-D-glucosaminidase, chitobiosidase and endochitinase) (Tronsmo and Harman, 1993) [72,73]. This enzyme is also classified as ß-N-acetyl-D-hexosaminide N-acetylhexosaminohydrolase because it acts on glucosides, galactosides, and several oligosaccharides (Webb, 1984) [73]. N-acetyl-ß-D-glucosaminidase (ß-glucosaminidase) is the enzyme that catalyses the hydrolysis of N-acetyl-ß-D-glucosamine residues from the terminal nonreducing ends of chitooligosaccharides [73]. This hydrolysis is important in carbon (C) and nitrogen (N) cycling in soils because it participates in the processes whereby chitin converts to amino sugars, which are major sources of mineralisable N in soils [72].

Bacteria are major mediators of chitin degradation in nature. In soil systems, chitin hydrolysis rates correlate with bacterial abundance (Kielak et al., 2013). However,

depending on temperature, pH, or the successional stage of the degradation process, fungi may also be quantitatively important agents of chitin degradation (Gooday, 1990; Hallmann et al., 1999; Manucharova et al., 2011) [74]. In aquatic systems, *Cytophaga-Flavobacteria* is known to profit from chitin addition and has been detected in a dense cluster on chitinous particles where they also assimilate chitin hydrolysis products (Cottrell and Kirchman, 2000; Beier and Bertilsson, 2011). This suggests a central role of *Cytophaga-Flavobacteria* in aquatic chitin degradation, where they also benefit from this material as a substrate [74].

The semi-quantitative API ZYM technique only allows the measurement of bacterial strains' potential enzymatic activity which is never identical to bacterial activity in natural environments [75]. API ZYM and API Coryne provide initial information about a strain's potential for in-depth exploration. In addition to these semi-quantitative tests, chemotaxonomy–in the form of a chemical biomarker approach by profiling cell envelope constituents such as phospholipids and fatty acids–is widely used in polyphasic taxonomy for the in-situ characterisation of microbial communities in natural environments [76].

3.5.3 Chemotaxonomic analysis

Menaquinone analysis was performed through the HPLC and HPLC-HRESIMS methods, where the menaquinone peak of strain 1932KM was compared to the peak from the type strain DSM 28230^{T}. The strain 1932KM was cultivated from a myxovirescin medium after incubation for 5 days at 30°C. The menaquinone peak of strain 1932KM appears at a retention time of 10.637 min, while for the type strain DSM 28230^{T} (*Flammeovirga arenaria*), it was at 10.254 min. Also, HPLC-HRESIMS shows that MK-7 (Menaquinone-7) has been identified at mass M+H 649 m/z and 671 m/z.

The major respiratory quinone is MK-7, in line with all other genera of the family *Flammeovirgaceae* except *Cesiribacter,* for which MK-4 was the major respiratory quinone [77]. Isoprenoid quinones are lipid molecules present in all respiratory and photosynthetic microorganisms specie and exhibit marked structural variations

depending upon the microbial taxon [76]. In prokaryotic cells, quinones are located exclusively in their cytoplasmic, intracytoplasmic and thylakoid membranes [76]. Microbial quinones are categorised into two major structural classes, the naphthoquinones and the benzoquinones, which are represented by menaquinones (formerly vitamin K2) and ubiquinones (formerly coenzyme Q) [76].

In addition to quinones, chemical biomarker approaches that rely on profiling cells envelope constituents such as phospholipids and FAs are widely used for the in-situ characterisation of microbial communities in natural environments as well [76]. For FA analysis of strain 1932KM using the FAME method, the amounts of FA were calculated in percentages using the value obtained from the integrated signal. The major FAs of strain 1932KM with nearly 5% or more are $C_{16:1}$ isomer 2 (30.0 %), *iso*-$C_{15:0}$ (19.0%) $C_{17:0}$ 3-OH (18.7 %), $C_{16:0}$ 3-OH (7.6%), $C_{18:0}$ (6.1%) and $C_{16:0}$ (5.6%) (Tables 2). The predominant FAs of *Flexithrix dorotheae*, the next relative, are $C_{16:\,1w5c}$ and *iso*-$C_{15:0}$ [66] (Table 3.7).

The identification of FAs in this system is accomplished by measuring retention time, which is the time for a specific FA to pass through the GC column [78]. To identify the various fatty acids, a numbering scheme is used that starts at the carboxyl carbon–the number one or α (alpha) carbon–and ending at the terminal, or ω (omega), carbon; the carboxyl carbon is labelled C-1 [78]. The basic structure of fatty acids is a carbon skeleton usually containing at least 12 carbon atoms with a carboxyl group (-COOH) at one end and a methyl group (-CH3) at the other [78]. The number of carbon atoms (x) is given, followed by a colon (:) and the number of double bonds in that molecule (y). The double bond and any other modifications to the carbon backbone - methyl groups (-CH3), hydroxyl groups (-OH)–are indicated by the number of carbon atoms from the carboxyl end (z). For example, a fatty acid named 18:1 ω5c (x,y z) would contain 18 carbon atoms and a single double bond, the double bond between the 13C-14C bond. Since the double bond begins at 13C in the 18:1 ω5c example, there are five carbon atoms from 13C to the ω carbon. Therefore, the ω5c notation indicates that the double bond is located five carbon atoms from the ω carbon [78].

Table 3.7 Cellular fatty acid profile of strain 1932KM after cultivation in CY-medium. Fatty acids representing more than 5% of the total are marked in bold.

Fatty Acids	1932KM
$C_{14:0}$ 3-OH	1.37
$C_{14:0}$ 2-OH	0.37
$C_{14:0}$	0.63
***iso*-$C_{15:0}$**	**18.97**
$C_{15:1}$	0.61
***iso*-$C_{15:1}$**	2.28
***iso*-$C_{15:0}$ 3-OH**	4.48
$C_{15:0}$	0.75
$C_{16:1}$ isomer 1	0.26
$C_{16:1}$ isomer 2	**29.97**
$C_{16:2}$	0.20
***iso*-$C_{16:0}$ 2-OH**	0.11
$C_{16:0}$ 3-OH	**7.60**
$C_{16:0}$	**5.59**
$C_{17:1}$	0.12
***iso*-$C_{17:1}$**	0.52
$C_{17:0}$ 3-OH	**18.69**
$C_{18:2}\omega6,9$ all cis	0.65
$C_{18:1}\omega9$ cis	0.60
$C_{18:0}$	**6.06**
$C_{24:0}$ 2-OH	0.17

Taxa are distinguishable by the types of fatty acids produced and the relative concentrations of individual fatty acids. In addition to genotype, the fatty acid composition is strongly influenced by three primary environmental variables—growth substrate, incubation temperature and incubation time—but results are highly reproducible, consistent and conserved among different taxa when these factors are held constant [79]. The most stable and reproducible cellular fatty acid profile is achieved by carefully regulating the growth conditions; to minimise these variables, there is a specific temperature and growth medium for each library [80]. However, many isolates showing minor genetic differences or variant genotypes cannot be distinguished with these tests [79].

The role of phenotypic and genotypic analysis in a polyphasic taxonomy study is not limited to identification; they can also be used to screen for bioactivity and gene clusters of secondary metabolites. In addition to studying the enzymatic reaction

and chemotaxonomy of strain 1932KM, we also performed an antibiotic sensitivity test.

3.5.4 Test for resistance against antibiotics

A test for resistance against 13 antibiotics was performed on VY/2 agar at 30°C for five days, as Mohr et al. [27] described. The final concentration was set to 100 µg.mL^{-1} for ampicillin, 30 µg.mL^{-1} for chloramphenicol, 10 µg.mL^{-1} for oxytetracycline, and 50 µg.mL^{-1} for the remaining antibiotics. For reference, the strain was grown on VY/2 agar without antibiotics. The result was that strain 1932KM is sensitive to ampicillin (Figure 2a), cephalosporin (Figure 2b), hygromycin (Figure 2f), fusidic acid (Figure 2k), thiostrepton (Figure 2l), and chloramphenicol (Figure 2 m). Meanwhile, under trimethoprim (Figure 2g), polymyxin (Figure 2h), bacitracin (Figure 2i) and oxytetracycline (Figure 2j), strain 1932KM could barely grow. Growth comparable to that on the plate without antibiotics (Figure 2n) is possible on kanamycin (Figure 2c), gentamicin (Figure 2d) and spectinomycin (Figure 2e).

Ampicillin and cephalosporin are ß-lactam antibiotics that prevent peptidoglycan synthesis of strain 1932KM's cell wall. Glycan polysaccharide strands are linked by a crosslink that binds polypeptides to N-acetyl muramic acid in each polysaccharide strand. After bactoprenol, a membrane-bound acceptor transfers UDP-NAMpentapeptide and UDP-NAG from the cytoplasm to the cell membrane's outer sid, transglycosylation and transpeptidation reactions are catalysed by penicillin-binding proteins (PBPs) bound to the cell membrane as a DD-peptidases to construct peptidoglycan [81].

These antibiotics act by inhibiting a set of transpeptidase enzymes (also called penicillin-binding proteins) unbound to the cell membranes essential for the synthesis of the peptidoglycan layer of the bacterial cell wall [82]. Meanwhile, hygromycin is an aminoglycoside antibiotic that inhibits protein synthesis [81]. The same mechanism had been attested for chloramphenicol and thiostrepton [81].

Finally, fusidic acid acts as an inhibitor of the translocation of elongation factor G (EF-G) by preventing EF-G releases from the ribosome [81].

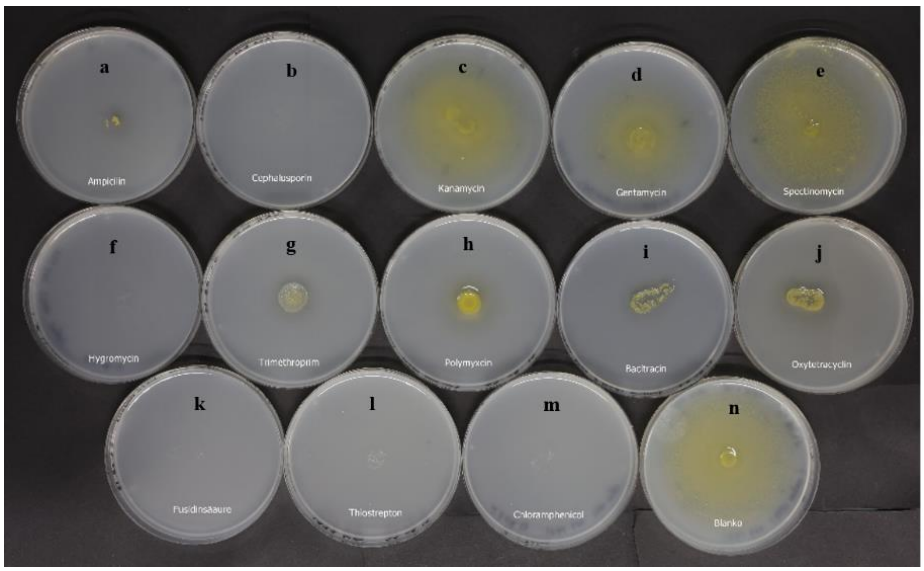


Figure 3.4 Antibiotic inhibition against strain 1932KM. (a)ampicillin, (b)cephalosporin, (c)kanamycin, (d)gentamicin, (e)spectinomycin, (f)hygromycin, (g)trimethoprim, (h)polymyxin, (i)bacitracin, (j)oxytetracycline, (k)fusidic acid, (l)thiostrepton, (m)chloramphenicol and (n)blank

In the potential antibiotics test against strain 1932KM, the strain showed resistance against kanamycin, gentamycin and spectinomycin. These antibiotics are members of the aminoglycoside group, which is the same group as chloramphenicol. While chloramphenicol is sensitive against strain 1932KM, these antibiotics display the opposite activity. If chloramphenicol is ineffective, it means the strain has the inactivator enzyme of Tn9 in the code for the mutated gene of ribosomal proteins [81]. The annotation of the genetic sequence for strain 1932KM by Prokka and the RAST analysis feature did not detect this gene either.

Kanamycin, gentamycin and spectinomycin as aminoglycoside antibiotics work when aminoglycosides bind to the aminoacyl-t-RNA recognition site (A-site) 16S rRNA that constitutes the 30S ribosomal subunit, leading to inhibition of polypeptide synthesis and subsequent cell death [83].

Resistance to aminoglycosides may occur through several mechanisms: (1) enzymatic modification and inactivation of the aminoglycosides, mediated by

aminoglycoside acetyltransferases, nucleotidyltransferases, or phosphotransferases and commonly observed across gram-positive and -negative bacteria; (2) increased efflux; (3) decreased permeability and (4) modifications of the 30S ribosomal subunit that interferes with binding of the aminoglycoside [83].

In *Mycobacterium tuberculosis*, aminoglycoside resistance is governed by point mutations to the 16S rRNA and the *rpsL* gene encoding the S12 protein [81]. Genetic analysis of strain 1932KM through Prokka and RAST shows that it contains *rpsL*, a gene that has been recognised as mycobacterium virulence operon involved in protein synthesis (SSU ribosomal proteins). This analysis explains the ineffectiveness of the aminoglycosides used in the potential antibiotics test. The result of the potential antibiotic test would be useful for investigating antibiotic resistance potential.

3.5.5 Gene cluster analysis

Microorganisms could produce antibiotics as secondary metabolites through enzymatic pathways known to be encoded in the biosynthetic gene cluster [84]. Some bioinformatics tools can analyse a gene cluster in a whole-genome sequence. In this thesis, AntiSMASH, NP.searcher, PubChem, SMART and Molinspiration web services have been used for this purpose.

AntiSMASH detected that strain 1932KM's genome contains trans-AT-PKS, NRPS, PKS-like (which has 17% similarity with polyketide), arylpoliene (30% similarity with polyketide), terpene (42% similarity with terpene), NRPS–T1PKS (8% similarity with NRP-Cyclic depsipeptide + polyketide: Iterative type I), and siderophore (20% similarity with other known clusters). The percentages of similarity indicate the comparisons between the gene clusters of strain 1932KM with the library provided by AntiSMASH.

NP.searcher is a tool to identify Type-I PKS, NRPS and hybrids of PKS/NRPS clusters in genomic data within the SMILES (Simplified Molecular Input Line Entry Specification) concerning the secondary metabolic products of the genome [85]. Type I PKSs is multifunctional peptides containing linearly arranged and

covalently fused domains [86]. The distinct domains work cooperatively and non-iteratively to catalyse carbon chain elongation and functional group regeneration [86].

Both AntiSMASH and NP.searcher have identified that strain 1932KM contains the NRPSs, PKSs, hybrid of NRPSs/PKSs, and terpenoid. Non-ribosomal peptide synthetases (NRPSs) and Polyketide synthases (PKSs) are large multifunctional enzymes of plants, bacteria and fungi that produce peptides and polyketides of broad structural and biological activity, respectively [86,87]. Meanwhile, hybrid synthetase is a representative of multifunctional enzymes containing an NRPS module and a PKS module on the same polypeptide chain, and it also yields predicted derivatives during fermentation [87]. Lastly, terpenoids have many important biological and physiological functions. In the late nineties, a novel non-mevalonate (non-MVA) pathway was discovered [88]. The compounds that result from one of these pathways would display activity against pathogens within its specific mechanisms.

Analysis with NP.searcher demonstrated that strain 1932KM's genome contains approximately 2 mixed modular NRPS/PKSs; 1 trans AT-PKS and 2 non-mevalonate terpenoid mep genes. These results are in the form of SMILES codes for cys-cys-mal and gly-mal-nrp-cys-cys domains.

The domain of mal (Malonamyl-CoA) is one of the starter units in Type II PKS and is mainly responsible for producing aromatic polyketides that are polycyclic compounds harbouring at least one aromatic ring [86]. They are an important type of natural product with antibacterial, anticancer and antiviral bioactivities, as exemplified by tetracyclines and anthracyclines [86].

Other than mal (Malonamyl-CoA), cys (Cysteine), gly (Glycine), and nrp (Non-ribosomal polyketide) are also starter units shown in the results from NP.searcher. They were then translated through the PubChem to obtain structural predictions. Figure 3.5 shows the predicted structures. The functional groups of the structures produced by this analysis are key in establishing specific interactions with target

proteins via the formation of hydrogen bonds and other weaker interactions, which also contribute to the metabolic stability and bioavailability of molecules [89].

(a)

(b)

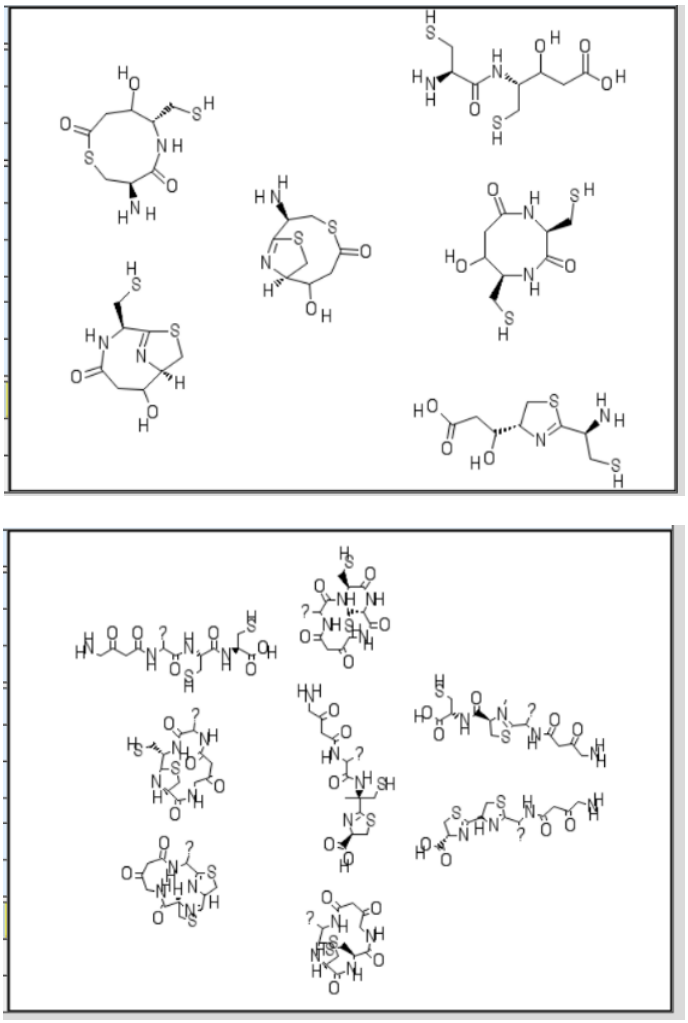


Figure 3.5 Predicted structures from strain 1932KM's genome. (a) detected cys-cys- mal and (b) gly-mal-nrp-cys-cys

Trans AT-PKSs is another gene cluster found by NP.searcher; it lies within the acyltransferase domain in the polyketide synthase (PKS) enzyme and is part of the NRPS/PKS domains. The explanation for this domain can be found in SMART.

AT-PKS is an enzyme similar to bacterial malonyl CoA-acyl carrier protein transacylase and eukaryotic FA synthase involved in FA biosynthesis. The group also includes polyketide synthases 6-methylsalicylic acid synthase (a multifunctional enzyme involved in the biosynthesis of patulin) and conidial green pigment synthase.

Regarding the metabolic pathways involving the proteins found in this domain, they include a protein that takes part in the biosynthesis of siderophore-group non-ribosomal peptides. This biosynthesis pathway produces some known compounds, e.g. Myxochelin A, Myxochelin B, Enterochelin, Bacilibactin, Vibriobactin, Pyochelin, Mycobactin, and Yersiniabactin. These compounds are siderophores that help the microbial system acquire iron. Blocking siderophore metabolism is one of the ways to inhibit bacterial growth. Siderophores also have applications in medicine as enhancers of potential antibiotics by taking on drug carriers' role.

The prediction of strain 1932KM's bioactivity potential can be performed with Molinspiration. This tool can predict bioactivity by its mechanisms. It gives a mechanism prediction score that demonstrates the possibility of evidence for the relevant activities. The larger the value, the higher probability that the molecule would be active. The data shows that the highest score is 2.00 points. The result for strain 1932KM from Moleinspiration includes the score for GPCR (G protein-coupled receptor) ligand (0.35), Ion channel modulator (0.31), Kinase inhibitor (−0.01), Nuclear receptor ligand (−0.17), Enzyme inhibitor (1.00), and Protease inhibitor (1.32). Molinspiration offers a prediction of bioactivity through various mechanisms, and in this respect, strain 1932KM displays high predicted values as an enzyme inhibitor and a protease inhibitor.

An enzyme inhibitor is a compound that binds to an enzyme and disrupts its activity. Such compounds could thus be potent poisons but could also be used as antibiotics and drugs to treat various diseases [90]. These compounds work by the mechanism of competitive or non-competitive inhibition [90]. Meanwhile, protease inhibitors are a class of drugs used to treat or prevent infection by viruses, including HIV and Hepatitis C [91]. Protease inhibitors prevent viral replication by inhibiting protease activity, e.g. HIV-1 protease, enzymes used by the viruses to cleave nascent proteins for the final assembly of new virions. Protease inhibitors have been developed or are presently undergoing testing for treating various viruses, e.g. HIV/AIDS antiretroviral protease inhibitors (saquinavir, ritonavir, indinavir, nelfinavir, amprenavir, etc.) [91].

Proteases (also termed peptidase or proteinase or proteolytic enzymes) are enzymes that hydrolyse the amide bonds of the peptide units that link amino acids in the polypeptides chain that form proteins [91,92]. They account for almost 6% of proteins in the human genome and constitute 1–5% of the genomes of bacteria, viruses and other infectious organisms [92]. Proteases are currently classified into four major groups according to the character of their catalytic active site and condition of action, i.e. serine, cysteine, aspartate, and metalloproteases [91]. The main intracellular proteases present in pathogenic bacteria, both Gram-positive and Gram-negative, include the ATP-dependent Clp, Lon, FtsH proteases, high-temperature requirement A (HtrA) protease, and the prokaryotic proteasome [92]. HtrA, Clp and Lon are serine proteases [92]. Prokka showed that strain 1932KM contains some of the main intracellular proteases of pathogenic bacteria, i.e. ATP-dependent Clp, Lon, and FtsH proteases.

3.5.6 Bioassay for antiinfectives

Screening for antiinfectives was performed on a methanol extract of strain 1932KM, which was fermented in 10 media (namely A, Cy, Cy/H, E, H, Myxovirescin, P, Pol, S, and VY/2) at 30°C for 5 days by serial dilution test and involved 9 targeted pathogenic bacteria. The pathogens were *Escherichia coli*, *Chromobacterium violaceum, Pseudomonas aeruginosa*, *Staphylococcus aureus*, *Micrococcus luteus*, *Mucor hiemalis*, *candida albicans*, and *Bacillus substilis.* The results displayed some activity against *Staphylococcus aureus*, *Micrococcus luteus*, *Mucor hiemalis, Candida albicans,* and *Bacillus substilis.* Unfortunately, this activity was not found in more than three-fold dilutions. This shows that the potency of the antiinfectives from these extracts is weak. The analysis of the highest activity from the extracts was based on the MyxoBase library since it is a known compound.

An interesting phenomenon was observed in the activity of the extracts in this project. The activity often disappears after the extract has passed through several recultures. There are several possible explanations for this. The first is the presence of contaminant. This disturbs the nutrient supply and thus the growth of the strain,

all of which would cause a decrease in the production of the active compound. The second is mutations.

In general, active compounds originate in a gene cluster in the strain, which forms part of the strain's DNA chromosomes. Bacterial chromosomes are complex, dynamic, and unstable. Instability can result from point mutations or genome rearrangements such as deletions, duplications, amplifications, insertions, inversions, or translocations that can disrupt genes and change information contained in the genome [93]. Some of these mutations can be silent, while others can lead to phenotypic variation, evolution and speciation [93]. The latter kind of mutations could alter the function or expression of proteins [93].

Both gene cluster and bioactivity analysis showed a synchronous result on the bioactivity test. Starter units that we found in strain 1932KM, i.e. PKS, NRPS, and their hybrid, were found on the chromosome to produce a compound [94]. Instability in the chromosome that controls the amount of active compound produced can be explored further to enhance productivity. Gene cluster analysis upon strain 1932KM showed that this strain has a chance to produce a new compound.

3.6 Accession and deposition number

Strain 1932KM has been submitted to GenBank under the accession number MF741332. The Indonesian Culture Collections also gave this strain the deposition number of InaCC B1242. We proposed the strain with the name *Balibacter flavus* gen.nov.,sp.nov.

3.7 Conclusion

Strain 1932KM is a gliding bacterium with yellow, non-motile, strictly aerobic, rod-shaped cells, liquefies agar and microaerophilic. Strain 1932KM displays weak starch degradation on P agar, but no yeast degradation (cleared zone) could be observed on VY/2 agar. Agar liquefaction occurred on CY agar with 1.5% NaCl. Growth was possible on the surface of the puncture-channel. Shaped with tapering ends, 0.7 µm wide and 5.6 µm long. Catalase activity was negative. Growth was

possible between 18°C and 38°C, with optimum growth conditions for strain 1932KM being 38°C in a pH range of 5.0–8.0. The major FAs of strain 1932KM^T with nearly 5% or more are $C_{16:1}$ isomer 2 (30.0%), *iso*-$C_{15:0}$ (19.0%) $C_{17:0}$ 3-OH (18.7%), $C_{16:0}$ 3-OH (7.6%), $C_{18:0}$ (6.1%), and $C_{16:0}$ (5.6%).

The API ZYM test showed strong activities in alkaline phosphatase, leucine arylamidase, valine arylamidase, acid phosphatase, α-galactosidase, α-glucosidase, and N-acetyl-β-glucosaminidase. Meanwhile, weak activity was observed for esterase (C4), cystine arylamidase, trypsin, α-chymotrypsin, naphthol-AS-Bl-phosphohydrolase, β-galactosidase, and α-fucosidase. By comparison, in the API Coryne test, strain 1932KM showed activity in pyrazinamidase, alkaline phosphatase, β-galactosidase, α-glucosidase, and β-glucosidase. Menaquinone 7 was identified in strain 1932KM as well. Also, this strain is sensitive to ampicillin, cephalosporin, hygromycin, fusidic acid, thiostrepton, and chloramphenicol.

Analysis of the full sequence of the 16S rRNA gene showed that the strain is closest to the family *Flammeovirgaceae*. The nomenclature of the family *Flammeovirgaceae* places it in phylum *Bacteroidetes*, class *Cytophagia*, order *Cytophagales*. The full 16S gene sequence of strain 1932KM has a homology of 91.3% with the type strain *Rapidithrix thailandica* (DSM 103551^T; LC 191863.1) and 90.9% with the type strain *Flexithrix dorotheae* (DSM 6795^T; NR 040919.1).

Housekeeping genes (pgm, pyrG and rpoB) were used to analyse dissimilarities between strain 1932KM and the type strains DSM 103551^T and DSM 6795^T according to their 16S rRNA gene sequences. The analysis parameters were identity matrix, conserved regions and single nucleotides. The results for strain 1932KM and DSM 6795^T were a score of 0.333 for the identity matrix, 7 conserved regions and 122 single nucleotides. Meanwhile, the multiple alignment analysis for strains 1932KM and DSM 103551^T resulted in a score of 0.303 for the identity matrix, 4 conserved regions and 109 single nucleotides. We also performed a statistical analysis of Spearman Correlation for these strains, which showed that the strains have certain similarities but are significantly different. According to the correlation size, which fell in the low positive correlation range between 0.3 and 0.5, we

propose to categorise this strain not only as a novel species but as a novel genus altogether.

Genome analysis on strain 1932KM with Prokka gave a result of 87 contigs, 8,681,004 bases, 6.689 CDS, 3 rRNA, 2 repeat regions, 46 t-RNA and 1 t-mRNA. The OrthoANIu value of strain 1932KM against DSM 6795^{T} is 69.90% while the DDH value is 19.60%. The G+C content of strain 1932KM is 39.6%. Gene cluster analysis with bioinformatics tools showed that strain 1932KM contains the domains of polyketide, non-ribosomal peptide and their hybrid. In addition, this strain has potential bioactivity as an enzyme inhibitor and protease inhibitor.

Polyphasic taxonomy studies have concluded that the novelty level of strain 1932KM is that of a new genus. Strain 1932KM has been submitted to GenBank under the accession number MF741332. This strain has also obtained a deposition number from the Indonesian Culture Collections, namely InaCC B1242 with the proposed name of *Balibacter flavus* gen.nov.,sp.nov.

3.8 Summary

A taxonomic study upon an isolated microorganism isolate is intended to find similarities and dissimilarities between the unknown organism and the closest type strain identified by BLAST and phylogenetic tree. The isolation of the microorganism reveals its bioactivity and the active compounds it produces. In general, we summarise the steps taken during polyphasic taxonomy analysis upon strain 1932KM below, as a comparison between strain 1932KM and the closest known type strains:

a. Analysis of 16S rRNA gene sequences, i.e. study of homologies/similarities; phylogenetic reconstruction; analysis of identity matrix, conserved regions and single nucleotides using housekeeping genes as a reference; analysis of the correlation between a candidate new strain and the relevant type strains using statistical methods, etc.

b. Analysis of whole-genome sequences using bioinformatics tools, i.e. annotating assembly sequences; describing strain features including G+C content, gene clusters, ANI and DDH; visualising the genome, etc.
c. Phenotypic analysis: colour, cell size, cell morphology, salt tolerance, optimum pH, optimum growth temperature, agar liquefaction capacity, microaerophilic character, etc.
d. Chemotaxonomic analysis: phospholipids, FA analysis, quinone analysis, enzyme properties etc.
e. Screening of bioactivity against target microorganisms.
f. Identification, isolation and purification of active compounds.

3.9 Recommendation

Due to 1932KM's potential for producing enzyme and protease inhibitors, this strain deserves further study into the genes responsible for these substances. Strain 1932KM certainly deserves further exploration as a new genus.

3.10 Acknowledgements

I would like to thank Kathrin I. Mohr, who chose strain 1932KM. Thank you to Chandra Risdian, Birte Trunkwalter, Stephanie Schultz, Klaus Peter Conrad, Wera Collisi, Esther Surges, and Kathrin's colleagues for helping me with phenotypic analysis and Aileen Gollasch for helping me with the chemical analysis.
Thank you to Indriawati, Uthayakumar Muthukumarashamy, Andreas Bremges, Till Robin Lesker, and also the team in the bioinformatics course in Gießen-Germany for sharing their knowledge of bioinformatics analysis. To Frank Klawonn and Raimo Franke for sharing their knowledge in statistical analysis.

3.11 References

[1] M. Gillis, P. Vandamme, P. D. Vos, et al., "Polyphasic taxonomy," *Bergey's manual trust*, pp. 1–10, 2015.

[2] R. R. Colwell, "Polyphasic taxonomy of the genus *Vibrio*: numerical taxonomy of *Vibrio cholerae*, *Vibrio parahaemolyticus*, and related *Vibrio* species," *Journal of bacteriology*, vol. 104, no. 1, pp. 410–433, 1970.

[3] P. Vandamme, B. Pot, M. Gillis, et al., “Polyphasic taxonomy, a consensus approach to bacterial systematics,” *Microbiological reviews*, vol. 60, no. 2, pp. 407–438, 1996.

[4] S. Das, H. R. Dash, N. Mangwani, et al., “Understanding molecular identification and polyphasic taxonomic approaches for genetic relatedness and phylogenetic relationships of microorganisms,” *Journal of microbiological methods*, vol. 103, pp. 80–100, 2014.

[5] J. Yoon, N. Oku, S. Park, et al., “*Tunicatimonas pelagia* gen. nov., sp. nov., a novel representative of the family *Flammeovirgaceae* isolated from a sea anemone by the differential growth screening method,” *Antonie van Leeuwenhoek*, vol. 101, no. 1, pp. 133–140, 2012.

[6] V. Raina, T. Nayak, L. Ray, et al., *A polyphasic taxonomic approach for designation and description of novel microbial species*, in: A polyphasic taxonomic approach for microbial diversity in the genomic era, Elsevier, pp. 137-152, 2019.

[7] J. Chun, A. Oren, A. Ventosa, et al., “Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes,” *International journal of systematic and evolutionary microbiology*, vol. 68, no. 1, pp. 461–466, 2018.

[8] K. H. Schleifer, “Classification of Bacteria and Archaea: past, present and future,” *Systematic and applied microbiology*, vol. 32, no. 8, pp. 533–542, 2009.

[9] F. A. Rainey, *How to Describe New Species of Prokaryotes*, in Methods in microbiology, Elsevier, vol. 38, pp. 7-14, 2011.

[10] D. L. J. Quicke, *Tertiary level biology: Principles and techniques of contemporary taxonomy*, Blackie Academic & Professional, 1st edition, 1993.

[11] M. Dworkin, S. Falkow, E. Rosenberg, et al., *The prokaryotes: Proteobacteria: Delta and epsilon subclasses*, Springer New York, New York, NY, 3rd edition, vol. 7, 2006.

[12] R. L. Hahnke, J. P. Meier-Kolthoff, M. García-López, et al., “Genome-based taxonomic classification of *Bacteroidetes*,” *Frontiers in microbiology*, vol. 7, p. 2003, 2016.

[13] J. P. Euzeby, “List of bacterial names with standing in nomenclature: a folder available on the internet,” *International journal of systematic bacteriology*, vol. 47, no. 2, pp. 590–592, 1997.

[14] A. C. Parte, “LPSN-list of prokaryotic names with standing in nomenclature,” *Nucleic acids research*, vol. 42, pp. D613-616, 2014.

[15] A. C. Parte, “LPSN - List of prokaryotic names with standing in nomenclature (bacterio.net), 20 years on,” *International journal of systematic and evolutionary microbiology*, vol. 68, no. 6, pp. 1825–1829, 2018.

[16] A. C. Parte, J. S. Carbasse, J. P. Meier-Kolthoff, et al., “List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ,” *International journal of systematic and evolutionary microbiology*, vol. 70, no. 11, pp. 5607-5612, 2020.

[17] H. Reichenbach, *The Order Cytophagales*, in: The Prokaryote, Springer, New York, NY, pp. 549-590, 2006.

[18] J. Yoon, K. Adachi, S. Park, et al., “*Aureibacter tunicatorum* gen. nov., sp. nov., a marine bacterium isolated from a coral reef sea squirt, and description of *Flammeovirgaceae* fam. nov,” *International journal of systematic and evolutionary microbiology*, vol. 61, Pt 10, pp. 2342–2347, 2011.

[19] M. Tang, C. Chen, J. Li, et al., “*Fabivirga thermotolerans* gen. nov., sp. nov., a novel marine bacterium isolated from culture broth of a marine cyanobacterium,” *International journal of systematic and evolutionary microbiology*, vol. 66, no. 2, pp. 1095–1099, 2016.

[20] M. Haber, S. Shefer, A. Giordano, et al., “*Fulvitalea axinellae* gen. nov., sp. nov., a member of the family *Flammeovirgaceae* isolated from the Mediterranean sponge *Axinella verrucosa*,” *International journal of systematic and evolutionary microbiology*, vol. 63, Pt 5, pp. 1678–1683, 2013.

[21] J. Yoon, Y. Matsuo, H. Kasai, et al., “*Limibacter armeniacum* gen. nov., sp. nov., a novel representative of the family 'Flammeovirgaceae' isolated from marine sediment,” *International journal of systematic and evolutionary microbiology*, vol. 58, Pt 4, pp. 982–986, 2008.

[22] J. Yoon, S. Ishikawa, H. Kasai, et al., “*Perexilibacter aurantiacus* gen. nov., sp. nov., a novel member of the family 'Flammeovirgaceae' isolated from sediment,” *International journal of systematic and evolutionary microbiology*, vol. 57, Pt 5, pp. 964–968, 2007.

[23] J. Yoon, N. Oku, S. Park, et al., “*Porifericola rhodea* gen. nov., sp. nov., a new member of the phylum *Bacteroidetes* isolated by the bait-streaked agar technique,” *Antonie van Leeuwenhoek*, vol. 100, no. 1, pp. 145–153, 2011.

[24] P. R. Subbarayan and M. Sarkar, “A comparative study of variation in codon 33 of the rpoS gene in *Escherichia coli* K12 stocks: implications for the synthesis of sigma(s),” *Molecular genetics and genomics: MGG*, vol. 270, no. 6, pp. 533–538, 2004.

[25] T. Iwase, A. Tajima, S. Sugimoto, et al., “A simple assay for measuring catalase activity: a visual approach,” *Scientific reports*, vol. 3, p. 3081, 2013.

[26] H. E. Schellhorn and V. L. Stones, “Regulation of katF and katE in *Escherichia coli K-12* by weak acids,” *Journal of bacteriology*, vol. 174, no. 14, pp. 4769–4776, 1992.

[27] K. I. Mohr, R. O. Garcia, K. Gerth, et al., “*Sandaracinus amylolyticus* gen. nov., sp. nov., a starch-degrading soil myxobacterium, and description of

Sandaracinaceae fam. nov," *International journal of systematic and evolutionary microbiology*, vol. 62, Pt 5, pp. 1191–1198, 2012.

[28] D. E. Minnikin, A. G. O'Donnell, M. Goodfellow, et al., "An integrated procedure for the extraction of bacterial isoprenoid quinones and polar lipids," *Journal of microbiological methods*, vol. 2, no. 5, pp. 233–241, 1984.

[29] J. Wink, P. Schumann, E. Atasayar, et al., "'Streptomyces caelicus', an antibiotic-producing species of the genus *Streptomyces*, and *Streptomyces canchipurensis* Li et al. 2015 are later heterotypic synonyms of *Streptomyces muensis* Ningthoujam et al. 2014," *International journal of systematic and evolutionary microbiology*, vol. 67, no. 3, pp. 548–556, 2017.

[30] R. Garcia, D. Pistorius, M. Stadler, et al., "Fatty acid-related phylogeny of myxobacteria as an approach to discover polyunsaturated omega-3/6 fatty acids," *Journal of bacteriology*, vol. 193, no. 8, pp. 1930–1942, 2011.

[31] K. Gemperlein, S. Rachid, R. O. Garcia, et al., "Polyunsaturated fatty acid biosynthesis in myxobacteria: different PUFA synthases and their product diversity," *Chemical science*, vol. 5, no. 5, pp. 1733-1741, 2014.

[32] R. R. Wick, L. M. Judd, C. L. Gorrie, et al., "Unicycler: resolving bacterial genome assemblies from short and long sequencing reads," *PLoS computational biology*, vol. 13, no. 6, pp. 1-22, 2017.

[33] T. Hall, "BioEdit version 5.0.6," North Carolina State University, Department of Microbiology, pp. 1–192, 2001.

[34] S. Kumar, G. Stecher, and K. Tamura, "MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets," *Molecular biology and evolution*, vol. 33, no. 7, pp. 1870–1874, 2016.

[35] Hammer, *PAST: Palaentological statistics-reference manual*, University of Oslo, version 3.25, pp. 1-275, 2019.

[36] T. Seemann, "Prokka: rapid prokaryotic genome annotation," *Bioinformatics (Oxford, England)*, vol. 30, no. 14, pp. 2068–2069, 2014.

[37] R. K. Aziz, D. Bartels, A. A. Best, et al., "The RAST server: rapid annotations using subsystems technology," *BMC genomics*, vol. 9, no. 75, pp. 1-15, 2008.

[38] T. Brettin, J. J. Davis, T. Disz, et al., "RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes," *Scientific reports*, vol. 5, pp. 8365, 2015.

[39] R. Overbeek, R. Olson, G. D. Pusch, et al., "The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST)," *Nucleic acids research*, vol. 42, pp. D206-214, 2014.

[40] P. Stothard and D. S. Wishart, “Circular genome visualization and exploration using CGView,” *Bioinformatics (Oxford, England)*, vol. 21, no. 4, pp. 537–539, 2005.

[41] J. R. Grant, A. S. Arantes, and P. Stothard, “Comparing thousands of circular genomes using the CGView comparison tool,” *BMC genomics*, vol. 13, no. 202, pp. 1-8, 2012.

[42] A. Petkau, M. Stuart-Edwards, P. Stothard, et al., “Interactive microbial genome visualization with GView,” *Bioinformatics (Oxford, England)*, vol. 26, no. 24, pp. 3125–3126, 2010.

[43] D. Zhang, F. Gao, I. Jakovlić, et al., “PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies,” *Molecular ecology resources*, vol. 20, no. 1, pp. 348–355, 2020.

[44] S.-H. Yoon, S. Ha, J. Lim, et al., “A large-scale evaluation of algorithms to calculate average nucleotide identity,” *Antonie van Leeuwenhoek*, vol. 110, no. 10, pp. 1281–1286, 2017.

[45] K. Blin, S. Shaw, K. Steinke, et al., “AntiSMASH 5.0: updates to the secondary metabolite genome mining pipeline,” *Nucleic acids research*, vol. 47, pp. W81-W87, 2019.

[46] N. Saitou and M. Nei, “The Neighbor-joining method: a new method for reconstructing phylogenetic trees,” *Molecular biology and evolution*, vol. 4, no. 4, pp. 406–425, 1987.

[47] R. C. Edgar, “Search and clustering orders of magnitude faster than BLAST,” *Bioinformatics (Oxford, England)*, vol. 26, no. 19, pp. 2460–2461, 2010.

[48] J. P. Meier-Kolthoff, A. F. Auch, H.-P. Klenk et al., “Genome sequence-based species delimitation with confidence intervals and improved distance functions,” *BMC bioinformatics*, vol. 14, no. 60, pp. 1–14, 2013.

[49] P. Glaser, L. Frangeul, C. Buchrieser, et al., “Comparative genomics of *Listeria* species,” *Science*, vol. 294, no. 5543, pp. 849–852, 2001.

[50] J. A. Vázquez-Boland, G. Domínguez-Bernal, B. González-Zorn et al., “Pathogenicity islands and virulence evolution in *Listeria*,” *Microbes and infection*, vol. 3, no. 7, pp. 571–584, 2002.

[51] J. Salinas, M. Zerial, J. Filipski, et al., “Nonrandom distribution of MMTV proviral sequences in the mouse genome,” *Nucleic acids research*, vol. 15, no. 7, pp. 3009–3022, 1987.

[52] G. Bernardi, “The vertebrata genome: isochores and evolution,” *Molecular biology and evolution*, vol. 10, no. 1, pp. 186–204, 1993.

[53] K. E. Simmon, A. C. Croft, and C. A. Petti, “Application of SmartGene IDNS software to partial 16S rRNA gene sequences for a diverse group of bacteria

in a clinical laboratory," *Journal of clinical microbiology*, vol. 44, no. 12, pp. 4400–4406, 2006.

[54] P. Stothard, "The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences," *Biotechniques*, vol. 28, no. 6, pp. 1102–1104, 2000.

[55] D. Campbell, *Nonparametric statistics*, in Volkmar F.R. (eds) encyclopedia of autism spectrum disorders, Springer, New York, NY, pp. 82-139, 2013.

[56] T. W. Epps and K. J. Singleton, "An omnibus test for the two-sample problem using the empirical characteristic function," *Journal of statistical computation and simulation*, vol. 26, 3-4, pp. 177–203, 1986.

[57] S. J. Goerg and J. Kaiser, "Nonparametric testing of distributions—the Epps–Singleton two-sample test using the empirical characteristic function," *The stata journal: Promoting communications on statistics and stata*, vol. 9, no. 3, pp. 454–465, 2009.

[58] P. Dalgaard, *Introductory statistics with R*, Springer New York, New York, NY, 2nd edition, 2008.

[59] M. M. Mukaka, "Statistic corner: a guide to appropriate use of correlation coefficient in medical research," *Malawi medical journal*, vol. 24, no. 3, pp. 69–71, 2012.

[60] R. O. Garcia, H. Reichenbach, M. W. Ring, et al., "*Phaselicystis flava* gen. nov., sp. nov., an arachidonic acid-containing soil myxobacterium, and the description of *Phaselicystidaceae* fam. nov," *International journal of systematic and evolutionary microbiology*, vol. 59, Pt 6, pp. 1524–1530, 2009.

[61] P. Srisukchayakul, C. Suwanachart, Y. Sangnoi, et al., "*Rapidithrix thailandica* gen. nov., sp. nov., a marine gliding bacterium isolated from samples collected from the Andaman sea, along the southern coastline of Thailand," *International journal of systematic and evolutionary microbiology*, vol. 57, Pt 10, pp. 2275–2279, 2007.

[62] N. R. Krieg, J. T. Staley, D. R. Brown, et al., *Bergey's manual of systematic bacteriology*, Springer New York, New York, NY, 2nd edition, vol. 4, 2010.

[63] Y. Muramatsu, M. Takahashi, M. Kaneyasu, et al., "*Persicobacter psychrovividus* sp. nov., isolated from shellfish, and emended descriptions of the genus *Persicobacter* and *Persicobacter diffluens*," *International journal of systematic and evolutionary microbiology*, vol. 60, Pt 8, pp. 1735–1739, 2010.

[64] M. T. Madigan, J. M. Martinko, and J. Parker, *Brock Mikrobiologie*, Spektrum Akademischer Verlag, Gustav Fischer, Heidelberg, 2001.

[65] H. Reichenbach, *Genus III. Flexithrix Lewin 1970, 513AL*, in Bergey's Manual of Systematic Bacteriology, Williams & Wilkins, Baltimore, 3rd edition, 1989.

[66] S. Hosoya and A. Yokota, “Reclassification of *Flexibacter aggregans* (Lewin 1969) Leadbetter 1974 as a later heterotypic synonym of *Flexithrix dorotheae* Lewin 1970,” *International journal of systematic and evolutionary microbiology*, vol. 57, Pt 5, pp. 1086–1088, 2007.

[67] D. Martínez, M. J. Molina, J. Sánchez, et al., “API ZYM assay to evaluate enzyme fingerprinting and microbial functional diversity in relation to soil processes,” *Biology and fertility of soils*, vol. 52, no. 1, pp. 77–89, 2016.

[68] S. C. Groom, M. J. Hazlett, and P. B. Little, “An evaluation of the API ZYM system as a means of identifying *Haemophilus somnus* and related taxa,” *Canadian journal of veterinary research*, vol. 50, no. 2, pp. 238–244, 1986.

[69] A. Soto, J. Zapardiel, and F. Soriano, “Evaluation of API Coryne system for identifying coryneform bacteria,” *Journal of clinical pathology*, vol. 47, no. 8, pp. 756–759, 1994.

[70] M. W. Humble, A. King, and I. Phillips, “API ZYM: a simple rapid system for the detection of bacterial enzymes,” *Journal of clinical pathology*, vol. 30, no. 3, pp. 275–277, 1977.

[71] W.-Y. Jeng, N.-C. Wang, M.-H. Lin, et al., “Structural and functional analysis of three β-glucosidases from bacterium *Clostridium cellulovorans*, fungus *Trichoderma reesei* and termite *Neotermes koshunensis*,” *Journal of structural biology*, vol. 173, no. 1, pp. 46–56, 2011.

[72] M. Ekenler and M. A. Tabatabai, “β-Glucosaminidase activity as an index of nitrogen mineralization in soils,” *Communications in soil science and plant analysis*, vol. 37, 7-8, pp. 1081–1094, 2004.

[73] M. Ekenler and M. Tabatabai, “Effects of trace elements on β-glucosaminidase activity in soils,” *Soil biology and biochemistry*, vol. 34, no. 11, pp. 1829–1832, 2002.

[74] S. Beier and S. Bertilsson, “Bacterial chitin degradation-mechanisms and ecophysiological strategies,” *Frontiers in microbiology*, vol. 4, pp. 1–12, 2013.

[75] Z. J. Mudryk and B. Podgórska, “Enzymatic activity of bacterial strains isolated from marine beach sediments,” *Polish journal of environment studies*, vol. 15, no. 3, pp. 441–448, 2006.

[76] A. Hiraishi, “Isoprenoid quinones as biomarkers of microbial populations in the environment,” *Journal of bioscience and bioengineering*, vol. 88, no. 5, pp. 449–460, 1999.

[77] T. N. R. Srinivas, P. A. Kumar, S. Madhu, et al., “*Cesiribacter andamanensis gen. nov., sp. nov.*, isolated from a soil sample from a mud volcano,” *International journal of systematic and evolutionary microbiology*, vol. 61, Pt 7, pp. 1521–1527, 2011.

[78] N. S. Sekora, K. S. Lawrence, P. Agudelo, et al., “Using FAME analysis to compare, differentiate, and identify multiple *Nematode* species,” *Journal of nematology*, vol. 41, no. 3, pp. 163–172, 2009.

[79] R. P. Larkin and C. L. Groves, “Identification and characterization of isolates of *Phytophthora infestans* using fatty acid methyl ester (FAME) profiles,” *Plant disease*, vol. 87, no. 10, pp. 1233–1243, 2003.

[80] M. Sasser, “Bacterial identification by gas chromatographic analysis of fatty acid methyl esters (GC-FAME),” *Technical Note #101*, 2006.

[81] S. Kırmusaoğlu, N. Gareayaghi, and B. S. Kocazeybek, “Introductory chapter: the action mechanisms of antibiotics and antibiotic resistance,” *IntechOpen*, pp. 1–10, 2019.

[82] T. Palzkill, “Metallo-β-lactamase structure and function,” *Annals of the New York academy of sciences*, vol. 1277, pp. 91–104, 2013.

[83] Y. Doi, J. Wachino, and Y. Arakawa, “Aminoglycoside resistance: the emergence of acquired 16S Ribosomal RNA Methyltransferases,” *Infectious disease clinics of North America*, vol. 30, no. 2, pp. 523–537, 2016.

[84] M. H. Medema and et al., “Minimum information about a biosynthetic gene cluster,” *Nature chemical biology*, vol. 11, no. 9, pp. 625–631, 2015.

[85] M. H. T. Li, P. M. Ung, J. Zajkowski, et al., “Automated genome mining for natural products,” *BMC bioinformatics*, vol. 10, pp. 1–10, 2009.

[86] J. Wang, R. Zhang, X. Chen et al., “Biosynthesis of aromatic polyketides in microorganisms using type II polyketide synthases,” *Microbial cell factories*, vol. 19, no. 1, pp. 1–11, 2020.

[87] R. Finking and M. A. Marahiel, “Biosynthesis of nonribosomal peptides,” *Annual review of microbiology*, vol. 58, pp. 453–488, 2004.

[88] V. S. Dubey, R. Bhalla, and R. Luthra, “An overview of the non-mevalonate pathway for terpenoid biosynthesis in plants,” *Journal of biosciences*, vol. 28, no. 5, pp. 637–646, 2003.

[89] P. Ertl, E. Altmann, and J. M. McKenna, “The most common functional groups in bioactive molecules and how their popularity has evolved over time,” *Journal of medicinal chemistry*, vol. 63, no. 15, pp. 8408–8418, 2020.

[90] T. Robin, S. Reuveni, and M. Urbakh, “Single-molecule theory of enzymatic inhibition,” *Nature communications*, vol. 9, no. 1, pp. 1–9, 2018.

[91] P. N. Sable, “Protease inhibitors: a review,” *Indian drugs*, vol. 50, no. 2, pp. 5–19, 2013.

[92] A. A. Agbowuro, W. M. Huston, A. B. Gamble, et al., “Proteases and protease inhibitors in infectious diseases,” *Medicinal research reviews*, vol. 38, no. 4, pp. 1295–1331, 2018.

[93] E. Darmon and D. R. F. Leach, “Bacterial genome instability,” *Microbiology and molecular biology reviews: MMBR*, vol. 78, no. 1, pp. 1–39, 2014.

[94] B. Callahan, M. Thattai, and B. I. Shraiman, “Emergent gene order in a model of modular polyketide synthases,” *Proceeding of the national academy of sciences of the United States of America*, vol. 106, no. 46, pp. 19410–19415, 2009.

Chapter 4. Analysis of housekeeping genes and multilocus sequences for a phylogenetic study of a potential novel Indonesian strain in the family *Myxococcaceae*

Corallococcus and Myxococcus are the most common genera of myxobacteria within the family Myxococcaceae. The 16S rRNA gene sequence analysis as a gold standard for identification method has certain limitations to its interpretative capability. Therefore, we made use of housekeeping genes as the main form of analysis upon similarities between our strains and the related strains. The analysis included the reconstruction of a phylogenetic tree and its polymorphism. Bioinformatic tools were used to support these analyses.

4.1 Background

Myxobacteria are single-celled but social eubacterial predators abundant in cultivated top-soils around the earth [1,2]. Members of the genera *Corallococcus* and *Myxococcus* within the family *Myxococcaceae* are common myxobacteria, widely encountered in the wild, with the ability to produce secondary metabolites. More than 1500 gene duplications, representing 15% of the genome, have been identified in *Myxococcus xanthus* that are lineage-specific [1,2]. These findings strongly suggest that the duplicated and diverged genes enabled evolution of the complex signalling required for myxobacteria's multicellular lifestyle [2].

Due to their distribution, we only have a limited understanding of their relationship and evolutionary history in various parts of the world. We thus sought to perform an initial analysis of the evolution of the strains based on phylogenetic studies. Phylogenetic analysis based on the 16S rRNA gene sequences alone may provide only a limited understanding of their relationships and evolutionary history [3]. BLAST results for 16S rRNA gene sequences of different unknown strains often are related to the same reference. We assume that if the strain does not display 100% similarity to the type strain, the strains must differ by at least one nucleotide in their sequences.

Since Pauling and Zuckerkandl proposed using gene sequences as a molecular clock and Woese and Fox introduced the use of rRNA genes to decipher phylogenetic

relationships, the 16S rRNA gene has been established as the gold standard in bacterial phylogeny [4–7]. It has been the universal marker gene for basic evolutionary analysis of both culturable and unculturable bacteria, thanks to its evolutionary conservation [8]. However, 16S rRNA gene sequences give inconsistent information on homologies with type strains in the National Center of Biotechnology Information database at the species level. We need a core method to identify a species' degree of relatedness to the type strains and their evolutionary relationship to each other. Housekeeping genes have become the core method for this kind of study[4].

Housekeeping genes (HKGs) are ubiquitously expressed in all tissue and cell types. They tend to be highly conserved and evolved slower than other genes (mainly due to their roles in the maintenance of basic cellular functions) and are essential for a cell's existence[9,10]. Selecting housekeeping genes as candidate reference genes according to their perceived stability in previous studies might be a useful strategy to increase the chances of validating a specific reference gene for a given bacterial species [11]. A scheme based on one to six genes as a reference should make use of genes coding for functional proteins (housekeeping genes) because they are considered more stable with respect to genetic mutations [4,8,12].

Mutation is the main study parameter for relatedness and homology between the strains. Strains that share identical allelic profiles refer to the same sequence type, while strains that only share some alleles are related and refer to sequence complexes [8]. Multilocus sequence analysis/typing (MLSA/MLST) is a nucleotide sequence-based approach for the unambiguous characterisation of prokaryotes via the Internet. It works by directly characterising DNA sequence variations in a set of housekeeping genes and evaluating relationships between strains based on their unique allelic profiles or sequences [13]. The relationship between the strains is based on the degree of relatedness shown in the phylogenetic tree developed out of 16S rRNA and/or housekeeping gene sequences.

Phylogenetic analysis on multilocus sequences involves several steps: (i) selection of strains and housekeeping genes, (ii) generation of sequences (polymerase chain

reaction (PCR) amplification and DNA sequencing), (iii) data analysis to identify homologous sites within each gene and (iv) MLSA using concatenated sequences or allelic profiles (the latter is usually used for population genetics) [13]. Bioinformatic tools can be applied to the analysis of phylogenetic reconstruction.

4.2 Hypothesis

Myxcoccaceae is the dominant family in the myxobacteria population, easy to isolate and identify on the first step of isolation. However, this also means that it is quite challenging to discover a new species in this family. The 16S rRNA gene sequence–as a gold standard for identifying homologies between a query strain and a type strain from an online database–often produces an outlier result from the consensus in identifying new species. There is always a chance of dissimilarity between strains, even with a homology score close to 100%. As such, there is still a good chance to find novel subspecies within the family *Myxococcaceae*. The most closely related strain identified from phylogenetic reconstruction results will be used as the base for further analysis. Also, the reconstruction needs core genes to set the boundaries of the analysis. The phylogenetic tree's core genes will be chosen from myxobacteria genes commonly found in the whole genome sequences from the NCBI database. When the strains have genetic differences, this will show up as dissimilarities in their nucleotide compositions.

4.3 State of the art

Ade´kambi et al. concluded that multilocus sequence analysis using three housekeeping genes as the core gene set might represent the first consensus and valid approach for investigating bacterial identification, phylogeny and taxonomy [12]. The use of a single-copy gene for community analysis is an important milestone in microbial ecology. It could allow for the accurate measurement of diversity and phylogenetic relationships, avoiding a loss in phylogenetic resolution and biases in diversity measurements due to the presence of intragenomic heterogeneity [4]. Based on the different recommendations mentioned above, at least four to five genes can be used, although, as often stated, even more genes

should be considered [8]. This analysis will use 3 housekeeping genes as the core for analysing the phylogenetic tree, namely phosphoglucomutase (pgm), cytidine triphosphate synthase (PyrG), and the β subunit of bacterial RNA polymerase (rpoB).

A multiple-sequence alignment needs to include all the sequences under comparison and cannot be aligned by considering only the nucleotide sequences [8]. The coding frames also need to be considered, especially if single amino acid insertions or deletions occur within a few sequences that have to be compared [8]. The housekeeping genes can be considered a reference for multiple alignments. In this thesis, we would like to use the concatenation of 3 housekeeping genes that we use to amplify the strains.

4.4. Materials and Methods

4.4.1 Strains

Sixteen strains from the family *Myxococcaceae* were isolated from several regions in Indonesia. Nine of these strains were from the genus *Myxococcus,* while seven were of the genus *Corallococcus*. These strains were provided by the GINAICO project and were originally isolated from 2016 to 2018. The type species belong to the MISG laboratory in HZI Germany. In this analysis, we use seven distinct type strains. Table 4.1 lists these strains. Meanwhile, the *Myxococcaceae* family strains from several regions in Indonesia are in Table 4.2.

The isolation method was taken from Shimkets et al. with some adaptation. All of the strains were reactivated from −80°C cryo-conserves in 20 ml Cy/H medium. They were incubated by shaking at 160 rpm and 30°C for 5 days.

Table 4.1 Table of type strains within the family *Myxococcaceae*

Type Strains Code	Strain Name
sbco036^T	*Corallococcus coralloides*
cce167^T	*Corallococcus exiguus*
ccm8^T	*Corallococcus macrosporus*
Mxf^T	*Myxococcus fulvus*
mxs8^T	*Myxococcus stipitatus*
sbmx085^T	*Myxococcus xanthus*
DSM 2260^T	*Myxococcus virescens*

Table 4.2 Table of strains from the family *Myxococcaceae*

Strains	Samples Source	Sequence Length (bp)	Similarity (%)	BLAST Result
BGB2139RAST-03	West Java	1501	99.79	*Myxococcus stipitatus*
MxGP53TM	West Java	1503	99.13	*Myxococcus stipitatus*
Mx1965KR	Bali	1508	99.93	*Myxococcus fulvus*
Mx1809KM	Jakarta	1509	99.88	*Myxococcus fulvus*
MxGP43TM	West Java	1503	99.20	*Myxococcus fulvus*
MxGP1963KM	Bali	1503	99.73	*Myxococcus Macrosporus*
TP13RAWA-A	West Java	1515	98.93	*Corallococcus exiguus*
Mxx1929(1)KR	Bali	1502	99.73	*Myxococcus macrosporus*
MxMK7KM	Southeast Sulawesi	1507	99.60	*Myxococcus macrosporus*
Mx2133KM	West Java	1503	99.87	*Myxococcus macrosporus*
CCGP6KM	West Java	1504	99.67	*Corallococcus coralloides*
Cce1965KR	West Java	1506	100	*Corallococcus coralloides*
Soce1969-2KM	West Java	1506	99.73	*Corallococcus coralloides*
CcMK6KM	Southeast Sulawesi	1506	99.93	*Corallococcus coralloides*
3RB1-1	West Java	1510	100	*Corallococcus exiguus*
MyxoGP11TM	West Java	1452	99.72	*Corallococcus exiguus*

4.4.2 Housekeeping genes

We used three HKGs in this analysis, i.e. cytidine triphosphate synthetase (pyrG), phosphoglucomutase (pgm), and RNA polymerase β-subunit (rpoB). Our references for these housekeeping genes are Santos & Ochman and Chen [14,15]. The primers for gene coding are listed together with references in Table 4.3. The table uses the ambiguity codes Y and S. According to IUPAC, Y refers to Cytosine (C) or Thymine (T) while S can be used for Guanine (G) or Cytosine (C).

Table 4.3 Housekeeping genes for reference

Genes	Primers	5'-3'	Amplicon Size (bp)
rpoB	rpoB- F	GCG ATC AAG GAG CGC ATG AG	1257
	rpoB-R	CCA CGG CAT GAA CGC GAC	
pyrG	pyrG-F	GAY CCS TAC ATC AAY GTS GAY	435
	pyrG-R	GTG CTGS GTG GGC TTS GTC TT	
pgm	pgm-F	CAT CTC SCA CGC SAT CCT C	1106
	pgm-R	AAG CTC TCC GCG TAG ATY TTG TAG A	

4.4.3 DNA extraction, polymerase chain reaction, amplification, and sequencing

Two millilitres of culture from reactivated strains were centrifuged at 11.000 rpm. The cell mass pellet was extracted using the Invisorb Spin Plant Mini Kit (Stratec Molecular) following the manufacturer's instructions. The extracted DNA, serving as a template with universal bacterial primer F27, R518, and R1492/R1525, was amplified in PCR for PCR proficiency. Then, the PCR products were examined by electrophoresis on 0.8% agarose.

One microliter of the DNA, extracted with PCR JumpStart Ready Mix (Sigma-Aldrich), was used according to the manufacturer's instructions. Isolates and type strains of family *Myxococcaceae* were amplified with pgm, pyrG, and rpoB using touchdown PCR, including an initial denaturation step at 94°C for 2 min, followed by 10 cycles of denaturation at 94°C for 1 min, annealing for 1 min (−1°C/cycle) at the primer-pair-specific annealing temperature at 60°C, and extension at 72°C for 1 min followed by 21 cycles and then repeated with denaturation at 94°C for 1 min,

annealing at 50°C for 1 min, and extension at 72°C for 1 min, with a final extension of 72°C for 5 min. We used the touchdown method for the pgm, pyrG, and rpoB genes from all strains in this study. We checked the quality of purified PCR-products once more on 0.8% agarose gels. Purified PCR products were sequenced in the Genome Analytics working group (HZI, Braunschweig, Germany).

4.4.4 Phylogenetic tree analysis

The sequences were checked for quality and assembled with BioEdit. The results were then identified and compared to the public database on the NCBI (National Center for Biotechnology Information) website. Phylogenetic tree reconstruction for each housekeeping gene was performed using MEGA 7 (Molecular Evolutionary Genetics Analysis) with the Neighbor-Joining method and 1000 bootstrap. Meanwhile, we analysed the phylogenetic tree with BioEdit to assess the degree of relatedness between the strains. The DnaSP v6.12 (DNA Sequence Polymorphism) software was also used to analyse its polymorphisms. The concatenation of housekeeping genes as a reference for alignment was performed with Phylosuite.

4.5 Results and discussion

4.5.1 Phylogenetic tree analysis

We performed a phylogenetic tree reconstruction for the 16 strains in the family *Myxococcaceae* from Indonesian isolates and the closest type strains from the MISG Lab.'s collections. The reconstruction for Indonesian strains used the full-length sequence of 16S rRNA gene sequences and housekeeping genes as well. On the other hand, we only used housekeeping gene sequences for the type strains. The results of this sequencing are summarised in Table 4.4.

The three conserved loci (pgm, pyrG and rpoB) were concatenated with Phylosuite. Then the strains' amplicons were aligned with the Clustal*W* tool in BioEdit version 7.

Table 4.4 16S and housekeeping gene sequence profiles for (a) and (b) unknown strains and (c) type strains

Length (bp)	MxMK7KM	MxGP1963KM	Mx2133KM	Mx1929(1)KR	MxGP53TM	BGB2139RAST03	Mx1809KM	MxGP43TM	Mx1965KR
16S	1507	1503	1503	1502	1503	1501	1509	1503	1508
pgm	1055	1069	1078	1069	1055	1053	1069	1092	1065
pyrG	432	429	432	430	422	430	432	424	430
rpoB	836	912	872	932	873	831	960	873	888

(a)

Length (bp)	Soce1969-2KM	CcMK6KM	CCGP6KM	Cce1965KR	3RB1-1	MyxoGP11TM	TP13RAWA-A
16S	1506	1506	1504	1506	1510	1452	1515
pgm	1084	1092	1059	1069	1052	1023	1055
pyrG	430	431	431	428	427	427	619
rpoB	920	860	841	901	910	849	952

(b)

Length (bp)	sbm085^T	Mxf^T	mxs8^T	DSM 2260^T	sbco036^T	cce167^T	ccm8^T
pgm	1082	1075	-	1092	1059	1079	1078
pyrG	429	429	433	432	431	425	435
rpoB	863	926	834	891	886	881	879

(c)

We chose pgm, pyrG, and rpoB to be the reference genes for several reasons. The pgm gene is the most widely distributed evolutionarily, with representatives easily identified in bacteria, especially as a contributor to the cell wall and polysaccharide capsular production [16]. PyrG was chosen since the pyrimidine nucleotide CTP and its derivatives are essential to all living organisms. Pyrimidine bases and nucleosides (the nucleotides' transportable precursor) are often unavailable as exogenous nutrients [17]. The enzymatic steps of the pyrimidine nucleotide biosynthetic pathway are the same in all bacteria [17]. However, the genomic organisation of the genes that encoded the pyrimidine biosynthetic enzymes and the mechanisms controlling these genes' expression, vary greatly from gene to gene and across the phylogenetic spectrum [17]. The last–rpoB–is the beta subunit of DNA polymerase. Vos et al. (2012) explained that the rpoB gene offers various potential advantages over standard 16S rRNA gene-based approaches in some earlier literature [18]. One reason is that rpoB typically occurs in a single copy. Another is that the higher resolution rpoB marker can reveal molecular variation down to the population level. Thirdly, the genetic divergence of rpoB correlates better with overall genomic divergence and provides better bootstrap support for phylogenetic reconstruction. Fourth, since rpoB is a protein-encoding gene, the data generated from this marker is more readily interpreted in an evolutionary framework. Fifth, (pyro)sequencing error is an important confounding factor in microbial diversity

studies using 16S rRNA gene sequences. Given that rpoB is a single-copy, essential protein-encoding gene, sequence errors can be readily identified and removed if they introduce disruptions to the reading frame.

All these advantages promote using the housekeeping genes as a core combine to produce a consistent position for the test strain among the phylogenetic tree branches. Therefore, the phylogenetic tree reconstructions were built by reference to the concatenated housekeeping genes. Figure 4.1 presents the reconstructed phylogenetic tree for the test strains.

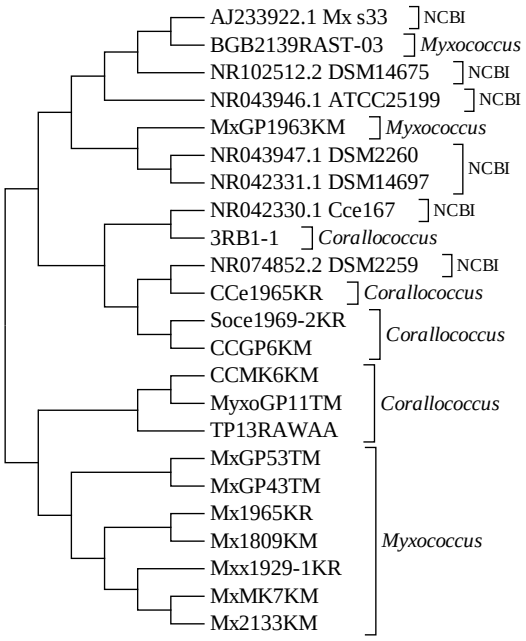
Phylogenetic analysis is important in clarifying the evolutionary pattern of multi-gene families and in understanding the process of adaptive evolution at the molecular level [19]. The reference is needed to align the taxa before phylogenetic reconstruction. Regarding Figures 4.1.a, 4.1.b, and 4.1.c, the MxGP1963KM strain showed up in inconsistent positions across the phylogenetic tree. However, in Figures 4.1.c and 4.1.d, this strain is consistently paired with strain Mxx1929-1KR. On the other hand, strains MxGP43TM and MxGP53TM appear consistently in the phylogenetic tree reconstructions.

These results indicate that the selection of candidate reference genes based on their perceived stability in previous studies might be a useful strategy to increase the chances of validating a specific reference gene for a given bacterial species [11]. The goal is to reconstruct a phylogenetic tree representing the evolutionary history of a population or a group of species [19].

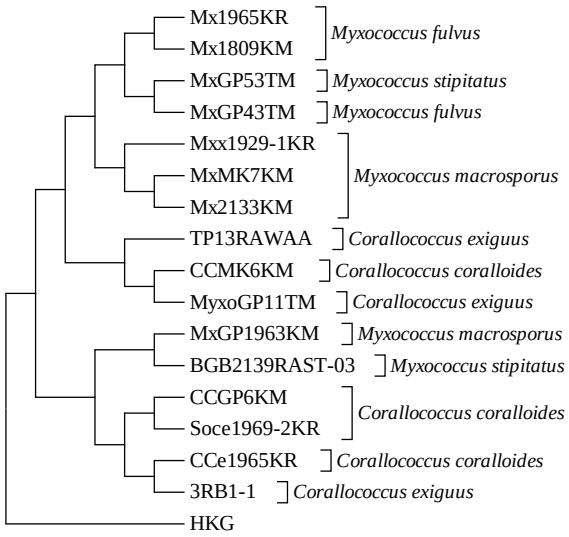
Amplifications of HKGs against strains and type species were then sequenced. After that, we analysed the sequences further by comparing them to the NCBI database. This would identify a type strain that possesses similar housekeeping genes. Then, all the results for each strain were concatenated using the Phylosuite software. Afterwards, the analysis proceeded into alignment processing with ClustalW, utilising the HKG concatenates as a reference. The distance between each pair of strains was used to perform phylogenetic tree reconstruction. Figure 4.1 shows the reconstruction results.

Multilocus analysis is the next step for the phylogenetic tree analysis. A critical point in studying MLSA is the selection of the housekeeping gene. Housekeeping genes coding for proteins with important functions should be considered because they are stable concerning rapid genetic modifications [8]. Also, strains that share identical allelic profiles refer to the same sequence type, whereas strains that only share some alleles are related and therefore refer to sequence complexes [8]. The referencing and alignment process plays an important part at this stage of the analysis.

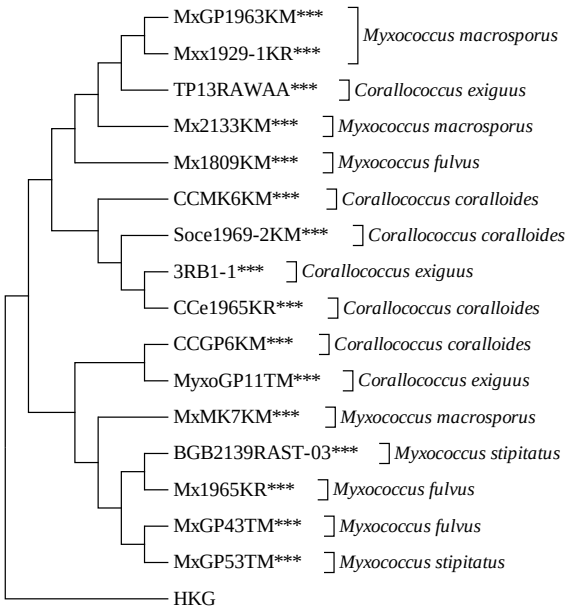
(a)



(b)



(c)



(d)

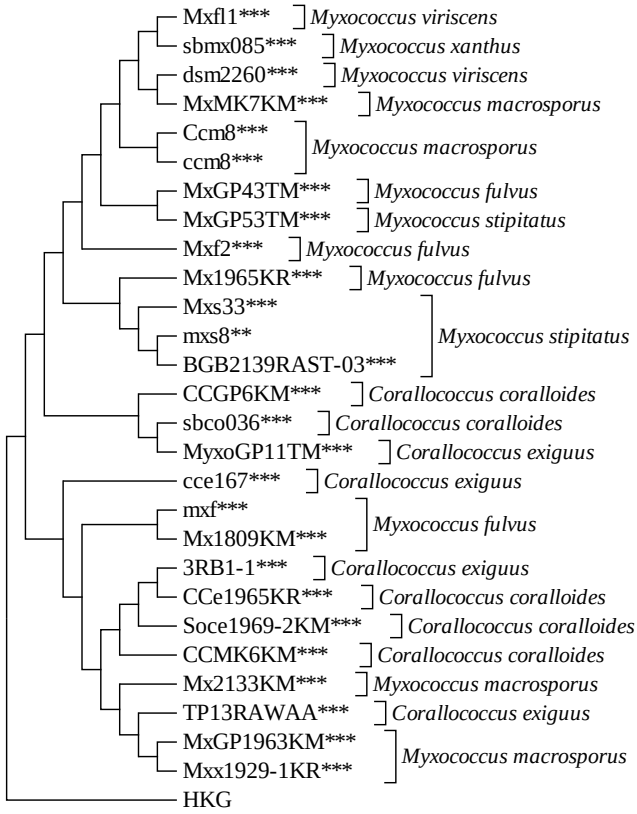


Figure 4.1 Phylogenetic trees reconstructed after 16S rRNA gene sequences and housekeeping genes. (a) 16S sequences with type strains as a reference, (b) 16S sequences with HKGs as a reference, (c) HKG sequences of unknown strains sequences with HKGs as a reference, (d) HKG sequences of unknown strains and type strains with HKGs as a reference.

Amplifications between the strains and the HKGs resulted in sequences with specific sequence lengths. Apart from these sequence lengths, nucleotide percentages can be calculated by MEGA7, and the results are presented in Table 4.5. Table 4.7 shows each strain's sequence length.

Table 4.5 Nucleotide frequencies of all strains from the phylogenetic tree in Figure 4.1.d

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
T(U) (%)	21.6	13.8	14.6	14.2	15.3	20.7	13.7	14.4	18.0	15.1	13.8	14.7	18.6	15.0	14.0
C (%)	21.0	34.2	33.5	34.5	33.6	30.6	35.0	34.7	33.4	34.2	33.6	34.3	34.1	33.7	33.8
A (%)	26.4	18.0	18.2	18.0	17.8	14.1	17.9	17.7	14.6	17.9	18.3	17.7	14.2	17.4	18.2
G (%)	31.0	33.9	33.7	33.2	33.2	34.6	33.4	33.2	34.0	32.9	34.3	33.3	33.2	33.8	34.0
Total	4224	1005	1004	1005	1005	425	879	891	926	834	886	863	910	831	841

	16	17	18	19	20	21	22	23	24	25	26	27	28	Average
T(U) (%)	18.1	18.6	17.9	14.6	18.1	17.9	14.0	14.7	14.5	17.9	13.3	18.3	17.8	16.8
C (%)	34.1	33.7	33.8	33.8	33.0	32.9	34.7	34.5	35.2	33.6	34.2	34.3	33.1	32.0
A (%)	14.0	14.0	14.4	17.8	13.5	14.4	17.6	17.8	17.5	13.5	18.1	14.2	13.9	17.9
G (%)	33.8	33.6	34.0	33.8	35.4	34.7	33.7	33.1	32.9	35.0	34.4	33.2	35.1	33.4
Total	860	901	960	888	912	872	873	873	836	932	849	920	948	1005.5

Note:

1. HKG
2. Ccm8^{T***}
3. Mxf2^{T***}
4. Mxfl1^{T***}
5. Mxs33^{T***}
6. cce167^{T***}
7. ccm8^{T***}
8. DSM 2260^{T***}
9. Mxf^{T***}
10. mxs8^{T**}
11. sbco036^{T***}
12. sbmx085^{T***}
13. 3RB1-1***
14. BGB2139RAST-03***
15. CCGP6KM***
16. CCMK6KM***
17. CCe1965KR***
18. Mx1809KM***
19. Mx1965KR***
20. MxGP1963KM***
21. Mx2133KM***
22. MxGP43TM***
23. MxGP53TM***
24. MxMK7KM***
25. Mxx1929-1KR***
26. MyxoGP11TM***
27. Soce1969-2KM***
28. TP13RAWAA***

HKG: Housekeeping genes
***: pgm, pyrG,and rpoB
**: pyrG and rpoB
T: Thymine
U: Uracil
A: Adenine
C: Cytosine
G: Guanine

Strains Ccm8^{T}, Mxf2^{T}, Mxfl1^{T} and Mxs33^{T} as the type strains were collected from NCBI as BLAST results. These were then processed to find similarities in amplicons between the unknown strains and HKGs in the database. In addition to nucleotide percentages and sequence lengths, we also calculated the distances between sequences in the same branch. MEGA7 analysis indicated that a pair of sequences with the closest distance will be in the same branch, as shown in Figure 4.1.d. Table 4.6 presents the profile distance results as an estimation of evolutionary divergence.

The analysis presented in Figure 4.1.d. involved 28 nucleotide sequences. All positions containing gaps and missing data were eliminated. There was a total of

394 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 [20].

Table 4.6 Estimates of evolutionary divergence between sequences of the phylogenetic tree in Figure 4.1.d

	sbmx085^{T}	**MxMK7KM**	**ccm8^{T}**	**MxGP53TM**	**BGB2139 RAST-03**
Mxfl1^{T}	0.018				
DSM2260^{T}		0.010			
Ccm8^{T}			0.000		
MxGP43TM				0.018	
mxs8^{T}					0.023

	MyxoGP11TM	**Mx1809KM**	**CCe1965KR**
sbco036^{T}	0.015		
Mxf^{T}		0.008	
3RB1-1			0.020

In the presence of polymorphic alleles, the divergence of genes sampled from different species should be longer than the species divergence [19]. The characterisation of this aspect can be accomplished through gel electrophoresis. This process must be performed before the amplicons are sequenced and the PCR must be done beforehand.

The ‘Touchdown’ method has been chosen for the PCR phase. Don et al. (1991) stated that it is used when a problem is encountered in the PCR amplification of target gene sequences, especially from complex genomes, namely the appearance of spurious smaller bands in the product spectrum [21]. Increasing the PCR’s annealing temperature may also solve the problem if we can presume the spurious interactions are sufficiently less stable than the specific (correct) ones due to degrees of sequence mismatch. This solution can minimise the need for time-consuming processes [21].

Sequencing and the subsequent analysis seek to prove the existence of polymorphism, starting with the analysis of the phylogenetic tree. Figure 4.1.d

showed that some strains are in the same branch, which means that they have some degree of shared homology.

The distance between strains in the same branch has the value closest to 0 (zero), as shown in Table 4.6. But even a zero value does not always indicate that the strains share 100% similarity or homology. For instance, strain ccm8^{T} as a type strain (from the MISG Lab strain collection) has a 100% homology to the type strain Ccm8^{T,} and we did not find any different alleles in the single nucleotide analysis. For the unknown strains, which have a 100% homology with the related strains, we do not expect to find differences in any single nucleotide between their sequences. However, strain MxGP1963KM and strain Mxx1929KR exhibit 99.78% similarity to each other, and we found two different alleles between them. Between strains MxGP1963KM and Mxx1929KR, one of them is possibly a subspecies of the other. However, this conclusion must be scrutinised further through a whole-genome sequence analysis more detailed comparisons to related type strain until we can obtain robust conclusions.

This way, the results could provide a reference for conclusions about strain novelty. The completed results of pairwise analyses on similarity and single nucleotides for Figure 4.1.d are shown in Table 4.7.

According to 16S rRNA gene sequence analysis, strain MxMK7KM has 99.60% homology with *Myxococcus macrosporus*; after amplification with the housekeeping genes, it has 99.40% relatedness to strain DSM 2260^{T} (*Myxococcus virescens*) with differences in 5 single nucleotides.

Table 4.7 Pairwise alignment analysis of sequences from the phylogenetic tree in Figure 4.1.d

Sequence 1	Sequence Length (bp)	Sequence 2	Sequence Length (bp)	Similarity (%)	Single Nucleotides
Mxfl1^T	1005	sbmx085^T	863	98.84	10
DSM 2260^T	891	MxMK7KM	836	99.40	5
Ccm8^T	1005	ccm8^T	879	100	-
MxGP43TM	873	MxGP53TM	873	96.79	28
mxs8^T	834	BGB2139RAST-03	831	97.69	19
sbco036^T	886	MyxoGP11TM	849	97.87	18
Mxf^T	926	Mx1809KM	960	98.18	11
3RB1-1	910	CCe1965KR	901	98.33	15
MxGP1963KM	912	Mxx1929-1KR	932	99.78	2

The similarity scores for other strains shown in Table 4.7 differ from previous similarity scores based on 16S rRNA gene sequences, which produced percentages ranging from 99% to 100%. The scores obtained from housekeeping genes displayed a much broader range from 97% to 100%, with a clearer picture of differences in single nucleotides. Also, the 100% similarity obtained between Ccm8^T and ccm8^T could reference other strains that display the same result. These strains do not differ by any nucleotides and their distance scores are 0 as well.

Estimating evolutionary divergence and the pairwise alignment analysis between sequences from phylogenetic tree is one method for finding a new species and/or subspecies. The novelty analysis was accomplished with whole-genome sequence analysis. Another phylogenetic tree analysis mode is studying conserved regions, and BioEdit can be applied to this purpose [22]. The analysis of conserved regions by BioEdit was based on multiple alignments. In this method, the concatenate of

HKGs is used as a reference, while the query sequences are the amplicons of HKGs from strains at the same position among the phylogenetic tree branches. The segment length parameter was adjusted until conserved regions were found, and the result was that we started with a length of 15 segments. The complete results for this are shown in Table 4.8.

A good phylogenetic marker displays some conserved regions, which are used as targets for universal probes. Primers and are useful to decipher phylogenetic relationships between distant organisms, as well as more variable regions, which are used as sites for genus- and family-level probes and primers and to differentiate closely related organisms [4]. Many housekeeping genes' repetitive natures are conserved between strains, and there are mechanisms to maintain sequence similarity between repeat copies within tandem arrays [23].

Table 4.8 Multiple alignment analysis of Figure 4.1.d by BioEdit

Reference	Sequence 1	Sequence 2	Conserved regions found	Regions	Consensus	Segment length
3 HKG	Mxfl1^T	sbmx085^T	1	1	2223 CGCGGCGCCGCTGGTGGGCAC 2243	21
	DSM 2260^T	MxMK7KM	1	1	2223 CGCGGCGCCGCTGGTGGGCAC 2243	21
	Ccm8^T	ccm8^T	2	1	2220 AC-CGCGGCGCCGCTGGTGGGCAC 2243	24
				2	2426 CAACCAGA--ACACCTGCCT 2445	20
	MxGP43TM	MxGP53TM	1	1	1968 CAGGCGAACGCGGA 1981	14
	mxs8^T	BGB2139RAST-03	2	1	1968 CAGGCGAACGCGGA 1981	14
				2	2185 CAACATGCAGCGTCA 2199	15
	sbco036^T	MyxoGP11TM	2	1	1785 CGGAAGGCCCGAACAT 1800	16
				2	2205 GTGCCGCTGCTGCG 2218	14
	mxf^T	Mx1809KM	4	1	1789 AGGCC 1793	5
				2	1976 CGCGG 1980	5
				3	2515 GACCG 2519	5
				4	2532 TGGCG 2536	5
	3RB1-1	CCe1965KR	2	1	1976 CGCGG 1980	5
				2	2532 TGGCG 2536	5
	MxGP1963KM	Mxx1929-1KR	4	1	1789 AGGCC 1793	5
				2	1976 CGCGG 1980	5
				3	2515 GACCG 2519	5
				4	2532 TGGCG 2536	5

4.5.2 Polymorphism analysis

Nei (2000) wrote that natural selection is an important factor in forming new species. Since the genetic variation of all morphological or physiological characters is ultimately controlled by variation at the protein or DNA level, it is important to identify the changes in protein or DNA sequences caused by natural selection [19].

Due to the importance of natural selection in the formation of new species, this study uses housekeeping genes to assess the possibility of discovering a new

subspecies. This analysis must look into polymorphism between one strain and its closest relative(s), using multiple alignments with concatenated HKGs as a reference. These sequences were concatenated to amplicons of strains with HKGs. These analyses also cover the various parameters of polymorphism such as sequence length, the number of sequences, the number of analyses, the net number of positions analysed, and the number of segregating sites. These methods are an approach to figure out evolutionary relationships by looking at mutations through statistical analysis.

Polymorphism between the strains was analysed with the DnaSP v6.12 software. DnaSP allows the analysis of DNA polymorphism data in specific functional regions [24]. It can compute many measures of the extent of DNA polymorphism anderform some common neutrality tests [24]. In studying strains located on the same branch, we used Multi-Domain Analysis as one of the features available in DnaSP. Table 4.9 presents the analytical results.

As we have shown previously in Table 4.7, strains Ccm8^{T} and ccm8^{T} display 100% shared similarity, implying both strains are identical. Meanwhile, according to the data presented in Table 4.9, the nucleotide diversity (Pi) for strains Ccm8^{T} - ccm8^{T} displays a value of 0.152, lower than other strains. This value is relevant in judging the similarity between the two strains in that the smaller the value, the higher the degree of similarity would be. Since DNA genetic information ultimately controls all morphological and physiological characters of organisms, any mutational changes in these characters are due to some change in DNA molecules [19]. Additionally, strain Ccm8^{T} is a reference strain which data are listed in the NCBI database. Its amplicon contains the similar HKGs like the type strain ccm8^{T}. Meanwhile, strain ccm8^{T} is the type strain deposited in the MISG's laboratory (original Reichenbach strain).

Table 4.9 Polymorphism analysis of strains from phylogenetic tree in Figure 4.1.d

Sequence 1	Sequence 2	Region	n	Sites	NetSites	S	Eta	Hap	Pi	AvNumDif	G+Ctot
Mxfl1^T	sbmx085^T	1-4311	3	4311	863	202	205	3	0.157	135.67	0.642
DSM 2260^T	MxMK7KM	1-4311	3	4311	836	195	195	3	0.156	130.00	0.645
Ccm8^T	ccm8^T	1-4311	3	4311	879	201	201	2	0.152	134.00	0.647
MxGP43TM	MxGP53TM	1-4311	3	4311	873	219	224	3	0.169	147.67	0.643
mxs8^T	BGB2139RAST-03	1-4311	3	4311	831	211	216	3	0.171	142.33	0.643
sbco036^T	MyxoGP11TM	1-4311	3	4311	847	207	207	3	0.163	138.00	0.649
mxf^T	Mx1809KM	1-4311	3	4311	874	606	608	3	0.463	404.67	0.642
3RB1-1	CCe1965KR	1-4311	3	4311	847	593	597	3	0.468	396.67	0.641
MxGP1963KM	Mxx1929-1KR	1-4311	3	4311	860	597	598	3	0.463	398.33	0.648

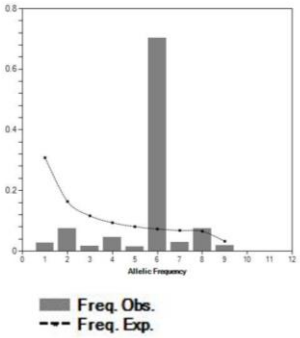
Note: *Region* is the length of sequences, *n* is the of number sequences, *Sites* is the number of analyses, *NetSites* is the net number of positions analysed, *S* is the number of segregating sites, *Eta* is the total number of mutation, *Hap* is the number of haplotypes, *Pi* is the nucleotide diversity, *AvNumDif (k)* is the average number of nucleotide differences, *G + Ctot* is the total GC content of the sequences [24]

The data input in multiple alignments (3 sequences) produced clear haplotype results for each pair of strains. A pair with a *Hap* number < 3 means the haplotypes in a sample of homologous DNA molecules are in a region with little or no mutation [25]. According to Pulido-Tamayo et al. (2015), the haplotypes used in these pairwise comparisons consist of base pairs corresponding to the size of bacterial sequences [26]. For this sequence analysis, haplotypes consist of certain base pairs corresponding to the size of a bacterial sequence, but the population (sequence length) typically contains fewer mutations [26]. This low mutation frequency implies an average distance between segregating sites [26].

The number of segregating sites (S) is the number of variable (i.e., polymorphic) positions in a sample of DNA sequences [27]. Nucleotide diversity (pi) signifies the average number of nucleotide differences per site (i.e. the probability that two random sequences are different at a given site) [27].

DnaSP shows (in tabular and graphic form) the distribution of the observed frequency spectrum (distribution of the allelic frequency in a site) and the expected values in a stable population, i.e., population with constant population size [24]. Figure 4.2 shows the distribution profile of allelic frequencies in paired sequences.

This figure explains population equilibrium and how mutations may push it into disequilibrium. This is shown by the fact that the spectrum of expected frequencies is not identical to that of observed frequencies.



S	Freq. Obs.	Freq. Exp.
1	0.026	0.308
2	0.073	0.164
3	0.015	0.117
4	0.046	0.093
5	0.014	0.081
6	0.703	0.073
7	0.028	0.068
8	0.075	0.065
9	0.019	0.032

Figure 4.2 Chart and table of polymorphic (segregating) sites of paired strains from Figure 4.1.d

Tajima **(**1989) developed a statistical method for testing the neutral mutation hypothesis by using the average number of nucleotide differences and the number of segregating sites [28]. *S* itself is not enough for estimating DNA polymorphism [29]. On the other hand, Tajima (1983) has shown under the neutral mutation model that the expectation and variance of the average number (*k*) of pairwise nucleotides' difference between the DNA sequences has a clear biological meaning [29]. Although the number of alleles (nucleomorphs) and the number of segregating sites is biologically clear, the average number of nucleotide differences can be used for measuring genetic variation within populations [30]. Furthermore, polymorphism analysis can help find an explanation for the differences between the strains.

4.6 Conclusion

Referring to the consensus value of a new species of 98.7%, despite the pairwise comparisons result between strains MxMK7KM:DSM 2260^{T} and Ccm8^{T}:ccm8^{T}, other strains still carry the possibility to be categorised as novel species. Furthermore, strain MxMK7KM could be a subspecies of strain DSM 2260^{T}. But these analyses must be corroborated further with a whole-genome sequence analysis to obtain a robust conclusion.

4.7 Summary

Housekeeping gene analysis is a method to find differences between bacteria by looking at their DNA segments. When we find a difference between bacteria at the DNA level, at least one nucleotide shows that the two bacteria involved must be classified differently. There are two possibilities that may result from the housekeeping gene method. First, we can find a new species or subspecies distinct from the type strain. Second, we may find a new subspecies of a type strain.

The use of housekeeping genes on this project has shown that the method can find clear similarities and dissimilarities between bacterial strains. This method can be recommended as a preliminary analysis before whole-genome sequencing in the effort to find a new species or subspecies.

4.8 Recommendation

Housekeeping gene analysis is a useful method that would be worth further development in taxonomy. Similar studies with amplicons of housekeeping genes from type strains that exist in online databases such as the NCBI's would be helpful in the future. Research in this field should continue with different type strains and different housekeeping genes.

4.9 Acknowledgements

Thank you to Joachim Wink and Kathrin I. Mohr, who entrusted this project to me. Also, I would like to thank Birte Trunkwalter and Senlie Octaviana for preparing and performing quality control upon the type strains for this project.

4.10 References

[1] D. Kaiser, M. Robinson, and L. Kroos, "Myxobacteria, polarity, and multicellular morphogenesis," *Cold spring harbor perspectives in biology*, vol. 2, no. 8, 1-26, 2010.

[2] B. S. Goldman, W. C. Nierman, D. Kaiser, et al., "Evolution of sensory complexity recorded in a myxobacterial genome," *Proceedings of the national academy of sciences of the United State of America*, vol. 103, no. 41, pp. 15200–15205, 2006.

[3] K.-S. Kim, K. S. Ko, M.-W. Chang et al., “Use of rpoB sequences for phylogenetic study of *Mycoplasma* species,” *FEMS microbiology letters*, vol. 226, no. 2, pp. 299–305, 2003.

[4] R. J. Case, Y. Boucher, I. Dahllöf, et al., “Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies,” *Applied and environmental microbiology*, vol. 73, no. 1, pp. 278–288, 2007.

[5] C. R. Woese and G. E. Fox, “Phylogenetic structure of the prokaryotic domain: the primary kingdoms,” *Proceedings of the national academy of sciences of the United State of America*, vol. 74, no. 11, 1977.

[6] C. R. Woese, “Bacterial evolution,” *Microbiological reviews*, vol. 51, no. 2, pp. 221–271, 1987.

[7] E. Zuckerkandl and L. Pauling, “Molecules as documents of evolutionary history,” *Journal of theoretical biology*, no. 8, pp. 357–366, 1965.

[8] S. P. Glaeser and P. Kämpfer, “Multilocus sequence analysis (MLSA) in prokaryotic taxonomy,” *Systematic and applied microbiology*, vol. 38, no. 4, pp. 237–245, 2015.

[9] H. S. Ong, “Comparative genomic analysis,” *Encyclopedia of Bioinformatics and Computational Biology*, vol. 3, pp. 425–431, 2019.

[10] J. Zhu, F. He, S. Hu et al., “Evolutionary steps of sex chromosomes are reflected in retrogenes,” *Trends in genetics: TIG*, vol. 24, no. 10, pp. 478–481, 2008.

[11] D. J. P. Rocha, C. S. Santos, and L. G. C. Pacheco, “Bacterial reference genes for gene expression studies by RT-qPCR: survey and analysis,” *Antonie van Leeuwenhoek*, vol. 108, no. 3, pp. 685–693, 2015.

[12] T. Adékambi, R. W. Butler, F. Hanrahan, et al., “Core gene set as the basis of multilocus sequence analysis of the subclass *Actinobacteridae*,” *PloS one*, vol. 6, no. 3, e14792, 2011.

[13] X. Rong and Y. Huang, *Methods in microbiology*: *multilocus sequence analysis*, Elsevier, United Kingdom, vol. 41, 2014.

[14] I.-C. K. Chen, B. Griesenauer, Y.-T. N. Yu, et al., “A recent evolutionary origin of a bacterial small RNA that controls multicellular fruiting body development,” *Molecular phylogenetics and evolution*, vol. 73, pp. 1–9, 2014.

[15] S. R. Santos and H. Ochman, “Identification and phylogenetic sorting of bacterial lineages with universally conserved genes and proteins,” *Environmental microbiology*, vol. 6, no. 7, pp. 754–759, 2004.

[16] K. M. Stiers, A. G. Muenks, and L. J. Beamer, “Biology, mechanism, and structure of enzymes in the α-D-phosphohexomutase superfamily,” *Advances in protein chemistry and structural biology*, vol. 109, pp. 265–304, 2017.

[17] C. L. Turnbough and R. L. Switzer, “Regulation of pyrimidine biosynthetic gene expression in bacteria: repression without repressors,” *Microbiology and molecular biology reviews: MMBR*, vol. 72, no. 2, pp. 266-300, 2008.

[18] M. Vos, C. Quince, A. S. Pijl, et al., “A comparison of rpoB and 16S rRNA as markers in pyrosequencing studies of bacterial diversity,” *PloS one*, vol. 7, no. 2, pp. 1-8, 2012.

[19] M. Nei and S. Kumar, *Molecular evolution and phylogenetics*, Oxford University Press, New York, 2000.

[20] S. Kumar, G. Stecher, and K. Tamura, “MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets,” *Molecular biology and evolution*, vol. 33, no. 7, pp. 1870–1874, 2016.

[21] R. H. Don, P. T. Cox, B. J. Wainwright, et al., “'Touchdown' PCR to circumvent spurious priming during gene amplification,” *Nucleic Acids Research*, vol. 19, no. 14, pp. 1, 1991.

[22] T. A. Hall, “BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT,” *Nucleic Acids Symposium Series*, no. 41, pp. 95–98, 1999.

[23] B. Andersson, “The *Trypanosoma cruzi* genome; conserved core genes and extremely variable surface molecule families,” *Research in microbiology*, vol. 162, no. 6, pp. 619–625, 2011.

[24] J. Rozas, “DnaSP: DNA sequence polymorphism - a manual,” Universitat de Barcelona, version 6.12, 2018.

[25] A. R. Templeton, *Human population genetics and genomics*: *Chapter 5: A backward view of genetic drift*, Elsevier, 2019.

[26] S. Pulido-Tamayo, A. Sánchez-Rodríguez, T. Swings et al., “Frequency-based haplotype reconstruction from deep sequencing data of bacterial populations,” *Nucleic Acids Research*, vol. 43, no. 16, e105, 2015.

[27] J. Rozas, *Chapter 17: DNA sequence polymorphism analysis using DnaSP*, in Bioinformatics for DNA sequence analysis, *Methods in molecular biology*, pp. 337–350, 2009.

[28] F. Tajima, “The effect of change in population size on DNA polymorphism,” *Genetics*, no. 123, pp. 597–601, 1989.

[29] F. Tajima, “Statistical method for testing the neutral mutation hypothesis by DNA polymorphism,” *Genetics*, no. 123, pp. 585–595, 1989.

[30] F.Tajima, “Evolutionary relationship of DNA sequences in finite populations,” *Genetics*, no. 105, pp. 437–460, 1983.

Chapter 5. Summary

Nature provides living things with everything they need through natural resources. The air, water and land each hold mysteries that have not been resolved. Humans, as living creatures themselves, bring problems to nature and nature itself provides the answer. However, humans must make an effort to find this answer since nature will not yield it readily. That is part of the mystery.

Every country in the world can contribute to solving this problem and Indonesia is no exception. However, this discovery cannot be accomplished on one's own. Instead, it requires multi-disciplinary cooperation with international researchers and GINAICO's project is an effort to bridge these conditions.

One of the aims of the GINAICO's project is the search for novel antimicrobial agents from the gliding bacteria with a focus on myxobacteria. Many gliding bacteria contain polyketide synthases and/or non-ribosomal peptide synthetase for secondary metabolites, which often show antimicrobial activities. As pathogens in this research, we have chosen *Staphylococcus aureus, Escherichia coli* and *Candida albicans*. All these pathogens have previously been found with resistance genes. *Methicillin-resistant Staphylococcus aureus, carbapenem-resistant Enterobacteriaceae* and *fluconazole-resistant Candida,* and went on to the further steps for analysis if we found a novel compound.

Antimicrobial resistance is one of the most serious problems faced by complex organisms on Earth. Many researchers have committed themselves to study microbes, and they have proven that microbes found in the wild have the potential to overcome humankind's problems with bacterial resistance. When an antimicrobial substance manages to enter a cell, many bacteria have another survival strategy to remove the substance. Research has to find strategies against the antimicrobial resistance problems today.

Choosing the appropriate research object (in this case, bacteria) is not easy. The bacterium must undergo screening to ensure that its distinct characteristics suit the problems we are trying to solve. Myxobacteria form the main object for this study due to their unique characteristics, which also offers opportunities to isolate other bacteria with similar morphology. This fact will be advantageous for the development of the taxonomy in general. Due to the high probability of success, rare bacterial species can be isolated through an accidental process.

According to the project, we have isolated 59 isolates of gliding bacteria, whereas 58 isolates are members of the myxobacteria, and 1 isolate is a non-myxobacterium. We identified that the non-myxobacterium could be a novel genus from the family *Flammeovirgaceae*. The study of the myxobacteria isolates in this research shows that most isolates have antimicrobial potential against *Staphylococcus aureus, Escherichia coli* and *Candida albicans*. The different activities are the results of different secondary metabolites.

Furthermore, handling bacteria that we have successfully isolated from the wild is often difficult. The conditions of its natural habitat would certainly differ from the environmental conditions in the laboratory. As a researcher, it is a challenge to bring the laboratory's environmental conditions as close as possible to the bacteria’s natural habitat. This includes the composition of the growing media, ambient temperature and pH. But that may not be enough; the bacterial community in the habitat may also become a factor worth considering when the extraction of secondary metabolites of the bacteria is one of the study's principal goals. This factor also influences the effectiveness of the silent genes.

The secondary metabolites produced by a bacterium are a self-defence system against changes in the environmental conditions around. They are also used for communication. However, it can also be a defence mechanism against attack by other bacteria. The rate of secondary metabolite production is tied to certain genetic mutations. These mutations can be studied at the molecular level and are also useful for taxonomic purposes. Mutations possessed by a bacterium can lead to the

identification of a new bacterial strain. This can be studied in more detail through phylogenetic tree reconstruction analysis.

The molecular biology of a bacterium plays an important part in its identification through polyphasic taxonomy. In addition to the molecular method, a conventional bacterial identification process includes morphological, physiological, enzymological and chemotaxonomic characteristics. When molecular taxonomy became the predominant trend, it tended to ignore existing methods of bacterial identification. This is not supposed to happen because the two methods complement each other. Therefore, we use the term ‘contemporary taxonomy’ to highlight the fact that there should not be any hard distinction between ‘conventional’ and ‘modern’ methods.

Contemporary taxonomy produces data that must be analysed. This analysis compares the new isolates strain to a reference strain. The development of bioinformatics research helps in the analysis of this molecular data. Bioinformatics has helped researchers in the field of taxonomy in identifying bacteria down to the single-nucleotide level. Bioinformatics can also make predictions about the secondary metabolites that a bacterium can produce and these metabolites’ potential biological activity. Also, bioinformatic analysis opens the door to the application of polyphasic taxonomy and to other fields, such as forensics.

When the information about a strain has been completely figured out, it does not represent the end but rather the beginning for further new research. Taxonomic identification is just the first step in exploring a strain’s potential. The research done in this project has pointed out the way for follow-up studies, including the critical point mentioned above. Biodiversity research stands to yield us more than just novel strains, compounds, or identification methods; it can also help us to understand the wisdom of nature.

Supplement 1. Media

1.1. Media for isolation, purification, and maintenance

1. Trace Element (Dworkin et.al, 2006)

$MnCl_2 \cdot 4H_2O$, Roth®	100 mg
$CoCl_2$, Roth®	20 mg
$CuSO_4$, Roth®	10 mg
$Na_2MoO_4 \cdot 2H_2O$, Merck®	10 mg
$ZnCl_2$, Merck®	20 mg
LiCl, Merck®	5 mg
$SnCl_2 \cdot 2H_2O$,Merck®	5 mg
H_3BO_3, Merck®	10 mg
KBr, Merck®	20 mg
KI, Merck®	20 mg
EDTA, Na-Fe^{3+} salt (tryhidrate), Fluka Chemie®	8 g
Water	1 litre

Sterilise by filtration and add 1 ml per litre of medium. The filter-sterilised solution is stable for months at room temperature because of its high EDTA content.

2. Water Agar (MISG-HZI)

$CaCl_2.2H_2O$, Roth®	1.5 g
$MgSO_4.7H_2O$, Roth®	1.5 g
HEPES, Roth®	11.8 g
Agar, Bacto Difco®	18.0 g

Water	1 litre

Check pH 7.2 and sterilise by autoclaving.

This medium also is able to combine with:

a. Combination 1:

Levamisole, Applichem Panreac®	100 mg
Soraphen A, produced by HZI	10 mg
Vitamin (B_{12}), Roth®	1 litre

b. Combination 2:

Cyclohexamide, Serva®	25 mg
Soraphen A, produced by HZI	10 mg
Vitamin (B_{12}), Roth®	1 litre

3. CY Agar (MISG-HZI)

Casitone, Bacto Difco®	3 g
Hefe extract (Marcor typ 9000), Roth®	1 g
$CaCl_2.2H_2O$, Roth®	1 g
HEPES, Roth®	11.9 g
Agar, Bacto Difco®	17g
Water	1 litre

Dissolve ingredients in distilled water and adjust to pH 7.2.

4. P agar

Pepton, Marcor S	2 g
Amylum, Cargil®	8 g

Probion MG069, Hoechst AG®	4 g
Yeast extract (Marcor typ 9000), Roth®	2 g
$CaCl_2$. $2H_2O$, Roth®	1 g
$MgSO_4$.4H2O, Roth®	1 g
100mM HEPES, Roth®	23.8 g/L
Fe-EDTA, Fluka Chemie®	8 mg/L
Agar, Bacto Difco®	18 g
Water	1 litre

Dissolve ingredients in distilled water and adjust to pH 7.5.

5. VY/2 agar

$CaCl_2$, Roth®	1 g
HEPES, Roth®	11.9 g
*Backerhef*e,	10 mL

(50% Hefe autoclaved or 1,5% Hefe extract)

Agar, Bacto Difco®	18 g

Dissolve ingredients in distilled water and adjust to pH 7.0.

After that added 1 mL vitamin B_{12} which is concentration of B_{12} = 0.5 g for 1 litre medium.

6. A medium

Glycerin (87% w/v), Roth®	0.4%
Soy flour (degreased), Cargil®	0.4%
Amylum (Cerestar), Cargil®	0.8%
Yeast extract (Marcor typ 9000), Roth®	0.2%

$CaCl_2.2H_2O$, Roth®	0.1%	
$MgSO4.7H_2O$, Roth®	0.1%	
50 mM HEPES (11.9 g/l), Roth®		
Fe-EDTA, Fluka Chemie®	8 mg/L	pH 7.4

7. CLF medium

Fructosa monohydrat, Cargil®	0.4%	
Glucose monohydrat, Cargil®	0.6%	
Skimmed milk	1%	
Yeast extract (Marcor typ 9000), Roth®	0.2%	
$CaCl_2.2H_2O$, Roth®	0.1%	
$MgSO_4.7H_2O$, Roth®	0.1%	
50 mM HEPES (11.9 g/l), Roth®		pH 7.0

8. E medium

Skimmed milk	0.4%	
Soy flour (degreased), Cargil®	0.4%	
Yeast extract (Marcor typ 9000), Roth®	0.2%	
Amylum (Cerestar), Cargil®	1%	
$MgSO_4.7H_2O$, Roth®	0.1%	
50 mM HEPES (11.9 g/l), Roth®		
Fe-EDTA, Fluka Chemie®	8 mg/L	pH 7.4
Glycerin (87% w/v), Roth®	0.5%	

9. H medium

Soy flour (degreased), Cargil®	0.4%	
Glucose monohydrat, Cargil®	0.2%	
Amylum (Cerestar), Cargil®	0.8%	
Yeast extract (Marcor typ 9000), Roth®	0.2%	
$CaCl_2.2H_2O$, Roth®	0.1%	
$MgSO4.4H_2O$, Roth®	0.1%	
50 mM HEPES (11.9 g/l), Roth®		
Fe-EDTA, Fluka Chemie®	8 mg/L	pH 7.2

10. M medium

Pepton (soy), Roth®	1%	
Maltose Monohydrat, Roth®	1%	
$CaCl_2.2H_2O$, Roth®	0.1%	
$MgSO4.4H_2O$, Roth®	0.1%	
50 mM HEPES (11.9 g/l), Roth®		
Fe-EDTA, Fluka Chemie®	8 mg/L	pH 7.2

11. Pol Medium

Amylum, Cargil®	0.30%	
Probion MG069, Hoechst AG®	0.30%	
$CaCl_2.2H_2O$, Roth®	0.05%	
$MgSO4.4H_2O$, Roth®	0.20%	
50 mM HEPES (11.9 g/l), Roth®		pH 7.2

12. S medium

Soy flour (degreased), Cargil®	0.4%	
Glucose monohydrat, Cargil®	0.2%	
Amylum (Cerestar), Cargil®	0.8%	
$CaCl_2.2H_2O$, Roth®	0.1%	
$MgSO4.4H_2$, Roth®	0.1%	
50 mM HEPES (11.9 g/l), Roth®		
Fe-EDTA, Fluka Chemie®	8 mg/L	pH 7.4

13. Myxovirescin Medium

Casein pepton, Marcor®	1%	
$CaCl_2.2H_2O$, Roth®	0.005%	
$MgSO4.4H_2$, Roth®	0.025%	
$CoCl_2$, Merck®	1 mg/L	
100 mM HEPES (23.8 g/l), Roth®		pH7.0

14. TAE-buffer (Tris-Acetat-EDTA; 1 L 50xTAE Buffer)

Tris, Sigma®	242 g
Acetic acid 100%, Roth®	57.1 mL
0.5M Na_2EDTA, Fluka Chemi®	100 mL

15. Mueller Hinton Broth, Roth® Art-Nr. bX927.1 ~ 21 g/L

Bovine infus	2 g/L	
Peptoe from Casein (Hydrolisate acid)	17.5 g/L	
Corn starch	1.5 g/L,	pH 7.4 ± 0.2

16. Myc Broth Medium

Phytone Peptone, BD®	10 g	
Glucose, Cerestar, Cargil®	10 g	
HEPES, Roth®	11.9 g	pH 7.0

17. Cy/H Medium

Cy medium	50%
H medium	50%

18. Artificial Sea Water

Coral Ocean Premium Quality Reef Salt
Free from synthetic additives and contains no nitrates, phosphates or silicates
Concentration 39 g/L
With Calcium 400–450 mg/L

19. Composition PCR solution

JumpStart™*Taq* ReadyMix™	12.50 µL
Riverse Primer 'x'	0.25 µL
Forward Primer 'x'	0.25 µL
PCR water	11.00 µL
DNA 'x'	1.00 µL

Note: Catalog product JumpStart™*Taq* ReadyMix™, Sigma-Aldrich. Catalog number P2893. This product contains:

20 mM Tris-HCl, pH 8.3

100 mM Kcl,

3 mM $MgCl_2$,

0.002% gelatin,

0.4 mM of each dNTP (dATP, dCTP, dGTP, TTP),

stabilisers,

unit/ µL Taq DNA Polymerase,

JumpStart Taq antibody.

Provided as 100 reactions and 400 reactions (50 µL reaction volume)

Supplement 2. List of in the phylum *Bacteroidetes* according to LPSN (List of Prokaryotic names with Standing in Nomenclature)

Class	Order	Family	Genus
***Bacteroidetes*, not assigned to class**	*Bacteroidetes*, not assigned to order	*Bacteroidetes*, not assigned to family	*Candidatus Amoebophilus, Candidatus Comitans, Karelsulcia, Magnispira, Toxothrix, Uzinura, Vallotia*
Bacteroidia	*Bacteroidales*	*Bacteroidales*, not assigned to family	*Candidatus Symbiothrix, Tammella, Vestibaculum*
		Bacteroidaceae	*Acetofilamentum, Acetothermus, Bacteroides, Desulfoarculus, Massilibacteroides, Phocaeicola*
		Balneicellaceae	*Balneicella*
		Barnesiellaceae	*Barnesiella, Coprobacter*
		Dysgonomonadaceae	*Dysgonomonas, Fermentimonas, Petrimonas, Proteiniphilum*
		Lentimicrobiaceae	*Lentimicrobium*
		Marinifilaceae	*Ancylomarina, Labilibaculum, Marinifilum*
		Marinilabiliaceae	*Alkaliflexus, Alkalitalea, Anaerophaga, Breznakibacter, Carboxylicivirga, Geofilum, Labilibacter, Mangroviflexus, Marinilabilia, Natronoflexus, Saccharicrinis, Thermophagus*
		Muribaculaceae	*Duncaniella, Muribaculum, Paramuribaculum*
		Odoribacteraceae	*Butyricimonas, Gabonibacter, Odoribacter*
		Paludibacteraceae	*Paludibacter*
		Porphyromonadaceae	*Falsiporphyromonas, Gabonia, Macellibacteroides, Microbacter, Oribaculum, Porphyromonas, Sanguibacteroides,*
		Prevotellaceae	*Alloprevotella, Capsularis, Hallella, Ihuprevotella, Marseilla, Massiliprevotella, Metaprevotella, Paraprevotella, Prevotella, Prevotellamassilia, Xylanibacter*
		Prolixibacteraceae	*Draconibacterium, Mangrovibacterium, Maribellus, Mariniphaga, Meniscus, Prolixibacter, Puteibacter, Roseimarinus, Sunxiuqinia, Tangfeifania*
		Rikenellaceae	*Alistipes, Anaerocella, Millionella, Mucinivorans, Rikenella, Tidjanibacter*
		Salinivirgaceae	*Salinivirga*
		Tannerellaceae	*Parabacteroides, Tannerella*
		Williamwhitmaniaceae	*Acetobacteroides, Williamwhitmania*
Chitinophagia	*Chitinophagales*	*Chitinophagaceae*	*Arachidicoccus, Arvibacter, Asinibacterium, Aurantisolimonas, Chitinophaga, Cnuella, Compostibacter, Crenotalea, Deminuibacter, Dinghuibacter, Edaphobaculum, Edaphocola, Ferruginibacter, Filimonas, Flaviaesturariibacter, Flavihumibacter, Flavipsychrobacter, Flavisolibacter, Flavitalea, Gynurincola, Haoranjiania, Heliimonas, Hydrobacter, Hydrotalea, Ilyomonas, Lacibacter, Mucibacter, Nemorella, Nemorincola, Niabella, Niastella, Niveitalea, Panacibacter, Paracnuella, Parafilimonas, Paraflavitalea, Parapseudoflavitalea, Parasediminibacterium, Parasegetibacter, Pseudobacter, Pseudoflavitalea, Puia, Rurimicrobium, Sediminibacterium, Segetibacter, Taibaiella. Terrimonas, Thermoflavifilum, Vibrionimonas*
Cytophagia	*Cytophagales*	*Bernardetiaceae*	*Bernardetia, Garritya, Hugenholtzia,*
		Catalimonadaceae	*Catalinimonas*
		Cesiribacteraceae	*Cesiribacter, Nafulsella*
		Cyclobacteriaceae	*Algoriphagus, "Anditalea", Aquiflexum, "Arthrospiribacter", Belliella, Cecembia, Chimaereicella, Cyclobacterium, Echinicola, Fontibacter, Hongiella, Indibacter, Lunatimonas, Mariniradius, Mongoliibacter, Mongoliicoccus, Mongoliitalea, Negadavirga, Nitritalea, Pleomorphovibrio, Rhodonellum, Shivajiella*
		Cytophagaceae	*Allopseudarcicella, Aquirufa, Arundinibacter, Chryseotalea, Cytophaga, Edaphorhabdus, Litoribacter, Rhodocytophaga, Sandaracinomonas, Sporocytophaga, "Tellurirhabdus"*
		[Cytophagales, not assigned to family]	*Luteivirga, "Taxeobacter"*
		Flammeovirgaceae	*"Algivirga", Flammeovirga, Flexithrix, Imperialibacter, Limibacter, Perexilibacter, "Porifericola", Rapidithrix, Sediminitomix, Tunicatimonas, Xanthovirga*
		Flexibacteraceae	*Flexibacter*
		Fulvivirgaceae	*Chryseolinea, Fulvivirga, Ohtaekwangia*

		Hymenobacteraceae	*Adhaeribacter, Botryobacter, Effluviibacter, Hymenobacter, Nibribacter, Pontibacter, Rufibacter*
		Marivirgaceae	*Marivirga*
		Microscillaceae	*Eisenibacter, Microscilla*
		Mooreiaceae	*Mooreia*
		Persicobacteraceae	*Aureibacter, Fulvitalea, Persicobacter*
		Raineyaceae	*Raineya*
		Reichenbachiellaceae	*Ekhidna, Marinoscillum, Reichenbachia, Reichenbachiella*
		Roseivirgaceae	*Fabibacter, Fabivirga, Marinicola, Roseivirga*
		Spirosomaceae	*Arcicella, Arcticibacterium, Arsenicibacter, Dyadobacter, Emticicia, Fibrella, Fibrisoma, Flectobacillus, Fluviimonas, Huanghella, Jiulongibacter, Lacihabitans, Larkinella, Leadbetterella, Nibrella, Persicitalea, Pseudarcicella, Ravibacter, Rhabdobacter, Rudanella, Runella, Siphonobacter, Spirosoma, Taeseokella, Telluribacter*
		Thermoflexibacteraceae	*Thermoflexibacter*
		Thermonemataceae	*Thermonema*
Flavobacteriia	*Flavobacteriales*	*Blattabacteriaceae*	*Blattabacterium*
		Crocinitomicaceae	*Brumimicrobium, Crocinitomix, Fluviicola, Lishizhenia, Putridiphycobacter, Salinirepens, Wandonia*
		Cryomorphaceae	*Cryomorpha, Luteibaculum, Phaeocystidibacter, Salibacter, Vicingus*
		Flavobacteriaceae	*Actibacter, Aequorivita, Aestuariibaculum, Aestuariicola, Aestuariimonas, Aestuariivivens, Algibacter, Algitalea, "Altuibacter", Amniculibacterium, Amoebinatus, Antarcticibacterium, Antarcticimonas, Aquaticitalea, Aquibacter Aquimarina, Arcticiflavibacter, Arenibacter, Arenitalea, "Ascidiaceibacter", Ascidiimonas, Aurantiacicella, Aurantibacter, Aurantivirga, "Aureibaculum", Aureicoccus, Aureisphaera, Aureitalea, Aureivirga, Bizionia, Capnocytophaga, Cellulophaga, Changchengzhania, "Citreibacter", Citreitalea, "Cochleicola", Coenonia, Confluentibacter, "Coraliitalea", Corallibacter, Costertonia, Croceibacter, Croceitalea, Croceivirga, Daejeonia, Dokdonia, Donghaeana, Eudoraea, Euzebyella, Faecalibacter, "Feifantangia", Flagellimonas, Flavicella, Flavimarina, Flaviramulus, Flavirhabdus, Flavivirga, Flavobacterium, Formosa, Frondibacter, Fulvibacter, Gaetbulibacter, Gaetbulimicrobium, Galbibacter, Gangjinia, Gelatiniphilus, Gelidibacter, Geojedonia, Gillisia, Gilvibacter, Gramella,. Haloflavibacter, Hanstruepera, Hemobacterium, Hoppeia, Hwangdonia, Hyunsoonleella, Ichthyenterobacterium, Imtechella, Jejudonia, Jejuia, Joostella, Kordia, Kriegella, Krokinobacter, Lacinutrix, Leeuwenhoekiella, Leptobacterium, Litoribaculum, Lutaonella, Lutibacter, Lutimonas, Mangrovimonas, Maribacter, Marinilexile, Marinirhabdus, "Marinitalea", Marinivirga, Maripseudobacter, Maritimimonas, Marixanthomonas, Meridianimaribacter, Mesoflavibacter, Mesohalobacter, Mesonia, Muricauda, Muriicola, Myroides, Namhaeicola, Neptunitalea, Nonlabens, Oceanihabitans, "Ochrovirga", Olleya, Paramesonia, Patiriisocius, Persicivirga, Pibocella, Planktosalinus, Polaribacter, Pontimicrobium, Pontirhabdus, Postechiella, Pricia, Pseudobizionia, Pseudofulvibacter, Pseudotenacibaculum, Pseudozobellia, Psychroflexus, Psychroserpens, "Pukyongia", Pustulibacterium, Robertkochia, Robiginitalea, Sabulilitoribacter, Salegentibacter, Salinimicrobium, Sandarakinotalea, Saonia, Sediminibacter, Sediminicola, Seonamhaeicola, Siansivirga, Sinomicrobium, Snuella, "Spodiobacter", Spongiibacterium, Spongiiferula, Spongiimicrobium, Spongiivirga, "Spongitalea", Stanierella, Stenothermobacter, Subsaxibacter, Subsaximicrobium, "Sufflavibacter", Sungkyunkwania, Taeania, Tamlana, Tenacibaculum, Ulvibacter, "Ulvibacterium", Urechidicola, "Vaginella", Vitellibacter, Wenyingzhuangia, Winogradskyella, Xanthomarina, Yeosuana, Zeaxanthinibacter, Zhouia*
		Ichthyobacteriaceae	*Ichthyobacterium*
		Schleiferiaceae	*Owenweeksia, Schleiferia, Thermaurantimonas*
		Weeksellaceae	*Algoriella, Apibacter, Bergeyella, Chishuiella, Chryseobacterium, Cloacibacterium, Cruoricaptor, Elizabethkingia, Empedobacter, Epilithonimonas, Kaistella, Moheibacter, Ornithobacterium, Planobacterium, Riemerella, Sejongia, Soonwooa, Spongiimonas, Wautersiella, Weeksella*
Saprospiria	*Saprospirales*	*Lewinellaceae*	*Flavilitoribacter, Haliscomenobacter, Lewinella, Phaeodactylibacter, Portibacter,*
		Saprospiraceae	*Aquirestis, Aureispira, "Candidatus Epifloribacter", Membranicola, "Rubidimonas", Saprospira*
Sphingobacteriia	*Sphingobacteriales*	*Crenotrichaceae*	*Crenothrix*
		Filobacteriaceae	*Filobacterium*
		Sphingobacteriaceae	*Albibacterium, Anseongella, Arcticibacter, Daejeonella, Mucilaginibacter, Nubsella, Olivibacter, Parapedobacter, "Pararcticibacter", Pedobacter, Pelobium, Pseudopedobacter, Pseudosphingobacterium, Solitalea, Sphingobacterium*
		Sphingobacteriales, not assigned to family	*"Candidatus Cardinium", "Candidatus Paenicardinium",*

Supplement 3. Reading profile API ZYM and API Coryne

3.1 Reading profile result API ZYM

No.	Enzyme Assayed for	Substrate	pH	Result	
				POSITIVE	NEGATIVE
1.	Control			Colorless or color of the sample if it has an intense coloration	
2.	Alkaline phosphatase	2-naphtyl phosphate	6.5	Violet	
3.	Esterase (C 4)	2-naphtyl butyrate	6.5	Violet	
4.	Esterase Lipase (C 8)	2-naphtyl caprylate	7.5	Violet	
5.	Lipase (C 14)	2-naphtyl myristate	-	Violet	
6.	Leucine arylamidase	L-leucyl-2-napthylamide	-	Orange	
7.	Valine arylamidase	L-valyl-2-napthylamide	-	Orange	
8.	Cystine arylamidase	L-cystyl-2-napthylamide	-	Orange	
9.	Trypsin	N-benzoyl-DL-arginine-2-naphthylamide	8.5	Orange	
10.	α-chymotrypsin	N-glutaryl-phenylalanine-2-naphtylamide	7.5	Orange	Colorless or
11.	Acid phosphatase	2-naphthyl phosphatase	5.4	Violet	Very pale
12.	Naphthol-AS-BI-phosphohydrolase	Naphthol-AS-BI-phosphate	-	Blue	yellow*
13.	α-galactosidase	6-Br-2-naphthyl-αD-galactopyranoside	-	Violet	
14.	β-galactosidase	2-naphthyl-βD-galactopyranoside	-	Violet	
15.	β-glucuronidase	Naphthol-AS-BI-βD-glucuronide	-	Blue	
16.	α-glucosidase	2- naphthyl-αD- glucopyranoside	-	Violet	
17.	β- glucosidase	6-Br-2-naphthyl-αD-glucopyranoside	-	Violet	
18.	N-acetyl-β-glucosaminidase	1-naphthyl-N-acetyl-βD-glucosaminide	-	Brown	
19.	α-mannosidase	6-Br-2-naphthyl-βD-mannopyranosidase	-	Violet	
20.	α-fucosidase	2-naphthyl-αL-fucopyranoside	-	Violet	

* Colourless or colour of the control if the strip has been exposed to an intense light source after the addition of the reagents. If the strip has not been exposed to intense light, a very pale yellow colour is obtained.

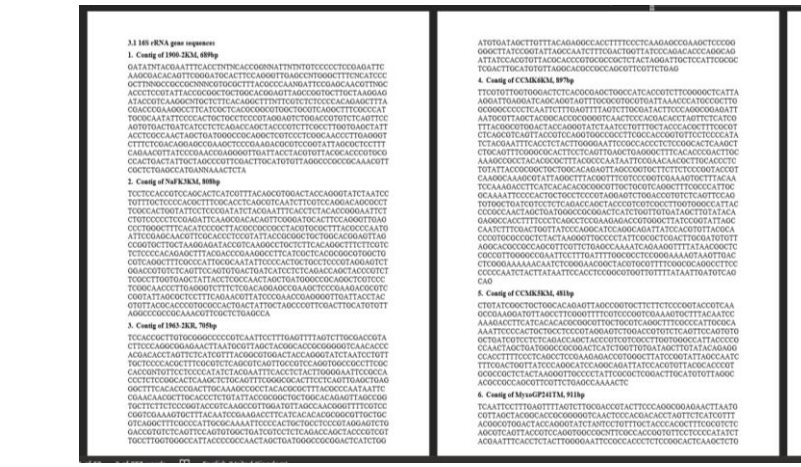
3.2 Profile reading API Coryne

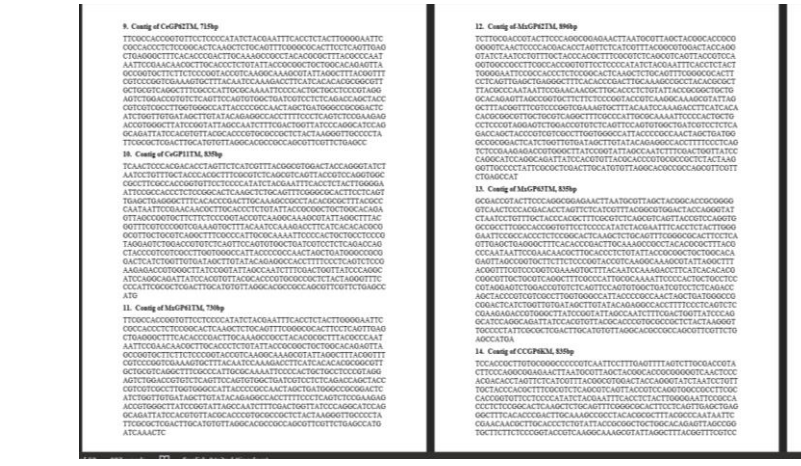
Tests	Active Ingredients	Qty (mg/cup.)	Reactions / Enzymes	Results	
				Negatives	Positives
NIT	Potassium nitrate	0.136	Reduction of NITrates	NIT 1+ NIT 2/10 min	
				Colorless Very pale pink	Dark pink Red
PYZ	Pyrazinecarboxamide	0.56	PYraZinamidase	PYZ / 10 min	
				Colorless Very pale brown Very pale orange	orange
PYRA	Pyroglutamic acid-β-naphthylamide	0.0256	PYRrolidonyl Arylamidase	ZYM A+ZYM B (PyrA→βNAG / 10 min	
				Colorless Pale orange	Orange
PAL	2-naphthyl-phosphate	0.0244	ALkaline Phosphatase	Colorless Beige–pale purple Pale orange	Purple
βGUR	Naphthol ASBI-glucuronic	0.0548	β-GlucURonidase	Colorless Pale grey Pale beige	Blue
βGAL	2-napthyl-βD-galactopyranoside	0.0312	β-GALactosidase	Colorless Beige-pale purple	Purple
αGLU	2-napthyl-αD-glucopyranoside	0.0308	α-GLUcosidase	Colorless Beige-pale purple Pale green	Purple
βNAG	1-naphthyl-N-acetyl-βD-glucosaminide	0.0348	N-Acetyl-β-Glucosaminidase	Colorless Beige-pale purple Pale brown Pale grey	Brown

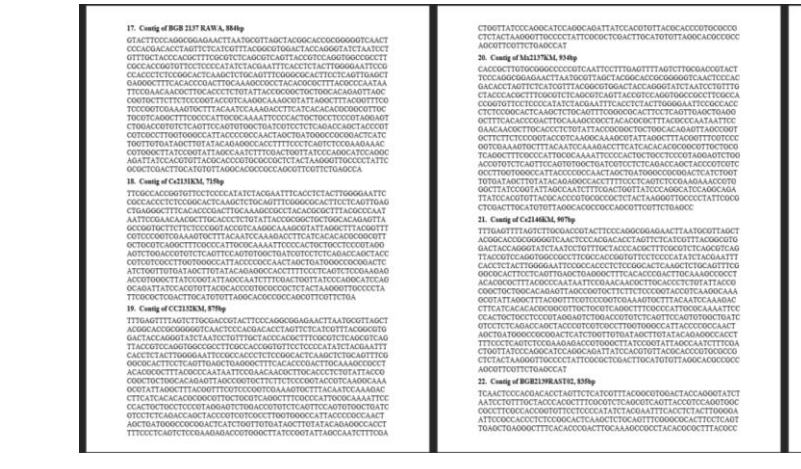
ESC	Esculin ferric citrate	0.546 0.078	β-glucosidase (ESCulin)	Colorless Grey	Black
URE	Urea	0.76	UREase	Yellow Orange	Red Pink
[GEL]	Gelatin (bovine origin)	0.6	Hydrolisis (GRLatin)	No diffusion of black pigment	Diffusion of black Pigment
0	Negative control	-	Fermentation		
GLU	D-glucose	1.56	Fermentation (GLUcose)		
RIB	D-ribose	1.4	Fermentation (RIBose)		
XYL	D-xylose	1.4	Fermentation (XYLose)		
MAN	D-mannitol	1.36	Fermentation (MANnitol)	Red Orange	Yellow Yellow-orange
MAL	D-maltose	1.4	Fermentation (MALtose)		
LAC	D-lactose (bovine origin)	1.4	Fermentation (LACtose)		
SAC	D-saccharose (sucrose)	1.32	Fermentation (SACcharosse)		
GLYG	Glycogen	1.28	Fermentation (GLYcoGen)		
CAT	(ESC or [GEL] test)	-	CATalase	H_2O_2 (3%) / 1 min	
				No bubbles	bubbles

- The quantity may be adjusted depending on the filter of the raw material used
- Certain cupules contain products of animal origin, notably pepton

Supplement 4. Fasta file of strains







CTACTAAGGGTTGCCCCTATTCGCGCTCGACTTGCATGTGTTAGGCACGCCGCCA
GCGTTCGTTCTGAGCCA

25. Contig of MxGP53TM, 832bp

TCAACTCCCACGACACCTAGTTCTCATCGTTTACGGCGTGGACTACCAGGGTATCT
AATCCTGTTTGCTCCCCACGCTTTCGCGTCTCAGCGTCAGTTACCGTCCAGGTGGC
CGCCTTCGCCACCGGTGTTCCTCCCCATATCTACGAATTTCACCTCTACTTGGGGA
ATTCCGCCACCCTCTCCGGCACTCAAGCACGACAGTTTCGGGCGCACTTCCTCAG
TTGAGCTGAGGGCTTTCACACCCGACTTGTCACGCCGCCTACACGCGCTTTACGC
CCAATAATTCCGAACAACGCTTGCACCCTCTGTATTACCGCGGCTGCTGGCACAG
AGTTAGCCGGTGCTTCTTCTCCCGGTACCGTCAAGCCGTAACGTATTGGGTTACG
GGTTTTCGTCCCGGTCGAAAGTGCTTTACAATCCGAAGACCTTCATCACACACGC
GGCGTTGCTGCGTCAGGCTTTCGCCCATTGCGCAAAATTCCCCACTGCTGCCTCCC
GTAGGAGTCTGGACCGTGTCTCAGTTCCAGTGTGGCTGATCGTCCTCTCAGACCA
GCTACCCGTCGTTGCCTTGGTGGGCCATTACCCCGCCAACTAGCTGATGGGCCGC
GGACTCATCTGGTTGTGATAGCTTGTATACAGAGGCCACCTTTTCCCTCAGTCTCC
GAAGAAACCGTGGGCTTATCCGGTATTAGCCAATCTTTCGACTGGTTATCCCAGG
CATCCAGGCAGATTATCCACGTGTTACGCACCCGTGCGCCGCTCTACTAAGGTTT
CCCTATTCGCGCTCGACTTGCATGTGTTAGGCACGCCGCCAGCGTTCGTTCTGAGC
C

26. Contig of MyxoGP31TM, 602bp

GACTTGTCACGCCGCCTACACGCGCTTTACGCCCAATAATTCCGAACAACGCTTG
CACCCTCTGTATTACCGCGGCTGCTGGCACAGAGTTAGCCGGTGCTTCTTCTCCCG
GTACCGTCAAGCCGTAGCGTGTTGGGCTACGGGTTTTCGTCCCGGTCGAAAGTGC
TTTACAATCCGAAGACCTTCATCACACACGCGGCGTTGCTGCGTCAGGCTTTCGC
CCATTGCGCAAAATTCCCCACTGCTGCCTCCCGTAGGAGTCTGGACCGTGTCTCA
GTTCCAGTGTGGCTGATCGTCCTCTCAGACCAGCTACCCGTCGTTGCCTTGGTGGG
CCATTACCCCGCCAACTAGCTGATGGGCCGCGGACTCATCTGGTTGTGATAGCTT
GTATACAGAGGCCACCTTTTCCCTCAGTCTCCGAAGAAACCGTGGGCTTATCCGG
TATTAGCCAATCTTTCGACTGGTTATCCCAGGCATCCAGGCAGATTATCCACGTGT
TACGCACCCGTGCGCCGCTCTACTAAGGGTTGCCCCTATTCGCGCTCGACTTGCAT
GTGTTAGGCACGCCGCCAGCGTTCGTTCTGAGCCATGATCAAACTCTA

27. Contig of CeGP512TM, 817bp

TAGTTCTCATCGTTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCC
ACGCTTTCGCGTCTCAGCGTCAGTTACCGTCCAGGTGGCCGCCTTCGCCACCGGT
GTTCCTCCCCATATCTACGAATTTCACCTCTACTTGGGGAATTCCGCCACCCTCTC
CGGCACTCAAGCACGACAGTTTCGGGCGCACTTCCTCAGTTGAGCTGAGGGCTTT
CACACCCGACTTGTCACGCCGCCTACACGCGCTTTACGCCCAATAATTCCGAACA
ACGCTTGCACCCTCTGTATTACCGCGGCTGCTGGCACAGAGTTAGCCGGTGCTTCT
TCTCCCGGTACCGTCAAGCCGGAGCGTGTTAGGCTCCGGGTTTTCGTCCCGGTCG
AAAGTGCTTTACAATCCGAAGACCTTCATCACACACGCGGCGTTGCTGCGTCAGG
CTTTCGCCCATTGCGCAAAATTCCCCACTGCTGCCTCCCGTAGGAGTCTGGACCGT
GTCTCAGTTCCAGTGTGGCTGATCGTCCTCTCAGACCAGCTACCCGTCGTCGCCTT
GGTGGGCCATTACCCCGCCAACTAGCTGATGGGCCGCGGACTCATCTGGTTGTGA
TAGCTTGTATACAGAGGCCACCTTTTCCCTCAGTCTCCGAAGAAACCGTGGGCTT
ATCCGGTATTAGCCAATCTTTCGACTGGTTATCCCAGGCATCCAGGCAGATTATCC

ACGTGTTACGCACCCGTGCGCCGCTCTACTAAGGTTTCCCTATTCGCGCTCGACTT
GCATGTGTTAGGCACGCCGCCAGCGTTCGTTCTGAGCCAT

28. Contig of MxGP43TM, 834bp

TCAACTCCCACGACACCTAGTTCTCATCGTTTACGGCGTGGACTACCAGGGTATCT
AATCCTGTTTGCTCCCCACGCTTTCGCGTCTCAGCGTCAGTTACCGTCCAGGTGGC
CGCCTTCGCCACCGGTGTTCCTCCCCATATCTACGAATTTCACCTCTACTTGGGGA
ATTCCGCCACCCTCTCCGGCACTCAAGCACGACAGTTTCGGGCGCACTTCCTCAG
TTGAGCTGAGGGCTTTCACACCCGACTTGTCACGCCGCCTACACGCGCTTTACGC
CCAATAATTCCGAACAACGCTTGCACCCTCTGTATTACCGCGGCTGCTGGCACAG
AGTTAGCCGGTGCTTCTTCTCCCGGTACCGTCAAGCCGTAGCGTGTTGGGCTACG
GGTTTTCGTCCCGGTCGAAAGTGCTTTACAATCCGAAGACCTTCATCACACACGC
GGCGTTGCTGCGTCAGGCTTTCGCCCATTGCGCAAAATTCCCCACTGCTGCCTCCC
GTAGGAGTCTGGACCGTGTCTCAGTTCCAGTGTGGCTGATCGTCCTCTCAGACCA
GCTACCCGTCGTTGCCTTGGTGGGCCATTACCCCGCCAACTAGCTGATGGGCCGC
GGACTCATCTGGTTGTGATAGCTTGTATACAGAGGCCACCTTTTCCCTCAGTCTCC
GAAGAAACCGTGGGCTTATCCGGTATTAGCCAATCTTTCGACTGGTTATCCCAGG
CATCCAGGCAGATTATCCACGTGTTACGCACCCGTGCGCCGCTCTACTAAGGGTT
GCCCCTATTCGCGCTCGACTTGCATGTGTTAGGCACGCCGCCAGCGTTCGTTCTGA
GCC

29. Contig of MxGP3KM, 884bp

GTACTTCCCAGGCGGAGAACTTAATGCGTTAGCTTCGGCACCGCGGGGGTCAACT
CCCACGACACCTAGTTCTCATCGTTTACGGCGTGGACTACCAGGGTATCTAATCCT
GTTTGCTCCCCACGCTTTCGCGTCTCAGCGTCAGTTACCGTCCAGGTGGCCGCCTT
CGCCACCGGTGTTCCTCCCCATATCTACGAATTTCACCTCTACTTGGGGAATTCCG
CCACCCTCTCCGGCACTCAAGCACGACAGTTTCGGGCGCACTTCCTCAGTTGAGC
TGAGGGCTTTCACACCCGACTTGTCACGCCGCCTACACGCGCTTTACGCCCAATA
ATTCCGAACAACGCTTGCACCCTCTGTATTACCGCGGCTGCTGGCACAGAGTTAG
CCGGTGCTTCTTCTCCCGGTACCGTCAAGCCGTAGCGTGTTGGGCTACGGGTTTTC
GTCCCGGTCGAAAGTGCTTTACAATCCGAAGACCTTCATCACACACGCGGCGTTG
CTGCGTCAGGCTTTCGCCCATTGCGCAAAATTCCCCACTGCTGCCTCCCGTAGGA
GTCTGGACCGTGTCTCAGTTCCAGTGTGGCTGATCGTCCTCTCAGACCAGCTACCC
GTCGTTGCCTTGGTGGGCCATTACCCCGCCAACTAGCTGATGGGCCGCGGACTCA
TCTGGTTGTGATAGCTTGTATACAGAGGCCACCTTTTCCCTCAGTCTCCGAAGAAA
CCGTGGGCTTATCCGGTATTAGCCAATCTTTCGACTGGTTATCCCAGGCATCCAGG
CAGATTATCCACGTGTTACGCACCCGTGCGCCGCTCTACTAAGGGTTGCCCCTATT
CGCGCTCGACTTGCATGTGTTAGGCACGCCGCCAGCGTTCGTTCTGAGCCA

30. Contig of MxGP1KM, 817bp

TAGTTCTCATCGTTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCC
ACGCTTTCGCGTCTCAGCGTCAGTTACCGTCCAGGTGGCCGCCTTCGCCACCGGT
GTTCCTCCCCATATCTACGAATTTCACCTCTACTTGGGGAATTCCGCCACCCTCTC
CGGCACTCAAGCACGACAGTTTCGGGCGCACTTCCTCAGTTGAGCTGAGGGCTTT
CACACCCGACTTGTCACGCCGCCTACACGCGCTTTACGCCCAATAATTCCGAACA
ACGCTTGCACCCTCTGTATTACCGCGGCTGCTGGCACAGAGTTAGCCGGTGCTTCT
TCTCCCGGTACCGTCAAGCCGTAGCGTGTTGGGCTACGGGTTTTCGTCCCGGTCG
AAAGTGCTTTACAATCCGAAGACCTTCATCACACACGCGGCGTTGCTGCGTCAGG

CTTTCGCCCATTGCGCAAAATTCCCCACTGCTGCCTCCCGTAGGAGTCTGGACCGT
GTCTCAGTTCCAGTGTGGCTGATCGTCCTCTCAGACCAGCTACCCGTCGTTGCCTT
GGTGGGCCATTACCCCGCCAACTAGCTGATGGGCCGCGGACTCATCTGGTTGTGA
TAGCTTGTATACAGAGGCCACCTTTTCCCTCAGTCTCCGAAGAAACCGTGGGCTT
ATCCGGTATTAGCCAATCTTTCGACTGGTTATCCCAGGCATCCAGGCAGATTATCC
ACGTGTTACGCACCCGTGCGCCGCTCTACTAAGGGTTGCCCCTATTCGCGCTCGA
CTTGCATGTGTTAGGCACGCCGCCAGCGTTCGTTCTGAGCC

31. Contig of MxGP5KM, 934bp

CACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTTAGTCTTGCGACCGTACT
TCCCAGGCGGAGAACTTAATGCGTTAGCTTCGGCACCGCGGGGGTCAACTCCCAC
GACACCTAGTTCTCATCGTTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTG
CTCCCCACGCTTTCGCGTCTCAGCGTCAGTTACCGTCCAGGTGGCCGCCTTCGCCA
CCGGTGTTCCTCCCCATATCTACGAATTTCACCTCTACTTGGGGAATTCCGCCACC
CTCTCCGGCACTCAAGCACGACAGTTTCGGGCGCACTTCCTCAGTTGAGCTGAGG
GCTTTCACACCCGACTTGTCACGCCGCCTACACGCGCTTTACGCCCAATAATTCCG
AACAACGCTTGCACCCTCTGTATTACCGCGGCTGCTGGCACAGAGTTAGCCGGTG
CTTCTTCTCCCGGTACCGTCAAGCCGTAGCGTGTTGGGCTACGGGTTTTCGTCCCG
GTCGAAAGTGCTTTACAATCCGAAGACCTTCATCACACACGCGGCGTTGCTGCGT
CAGGCTTTCGCCCATTGCGCAAAATTCCCCACTGCTGCCTCCCGTAGGAGTCTGG
ACCGTGTCTCAGTTCCAGTGTGGCTGATCGTCCTCTCAGACCAGCTACCCGTCGTT
GCCTTGGTGGGCCATTACCCCGCCAACTAGCTGATGGGCCGCGGACTCATCTGGT
TGTGATAGCTTGTATACAGAGGCCACCTTTTCCCTCAGTCTCCGAAGAAACCGTG
GGCTTATCCGGTATTAGCCAATCTTTCGACTGGTTATCCCAGGCATCCAGGCAGA
TTATCCACGTGTTACGCACCCGTGCGCCGCTCTACTAAGGGTTGCCCCTATTCGCG
CTCGACTTGCATGTGTTAGGCACGCCGCCAGCGTTCGTTCTGAGCC

32. Contig of BGB2139RAST03, 470bp

CCGTCAGCCGTAGCGTGTTGGGCTACGGGTTTTCGTCCCGGTCGAAAGTGCTTTA
CAATCCGAAGACCTTCATCACACACGCGGCGTTGCTGCGTCAGGCTTTCGCCCAT
TGCGCAAAATTCCCCACTGCTGCCTCCCGTAGGAGTCTGGACCGTGTCTCAGTTCC
AGTGTGGCTGATCGTCCTCTCAGACCAGCTACCCGTCGTTGCCTTGGTGGGCCATT
ACCCCGCCAACTAGCTGATGGGCCGCGGACTCATCTGGTTGTGATAGCTTGTATA
CAGAGGCCACCTTTTCCCTCAGTCTCCGAAGAAACCGTGGGCTTATCCGGTATTA
GCCAATCTTTCGACTGGTTATCCCAGGCATCCAGGCAGATTATCCACGTGTTACG
CACCCGTGCGCCGCTCTACTAAGGTTTCCCTATTCGCGCTCGACTTGCATGTGTTA
GGCACGCCGCCAGCGTTCGTTCTGAGC

33. Contig of Mx2134KM, 881bp

TGCAGTCGAGCGCGAATAGGGAAACCTTAGTAGAGCGGCGCACGGGTGCGTAAC
ACGTGGATAATCTGCCTGGATGCCTGGGATAACCAGTCGAAAGATTGGCTAATAC
CGGATAAGCCCACGGTTTCTTCGGAGACTGAGGGAAAAGGTGGCCTCTGTATACA
AGCTATCACAACCAGATGAGTCCGCGGCCCATCAGCTAGTTGGCGGGGTAATGGC
CCACCAAGGCGACGACGGGTAGCTGGTCTGAGAGGACGATCAGCCACACTGGAA
CTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTGCGCAAT
GGGCGAAAGCCTGACGCAGCAACGCCGCGTGTGTGATGAAGGTCTTCGGATTGT
AAAGCACTTTCGACCGGGACGAAAACCCGTAGCCTAACACGCTACGGCTTGACG
GTACCGGGAGAAGAAGCACCGGCTAACTCTGTGCCAGCAGCCGCGGTAATACAG

AGGGTGCAAGCGTTGTTCGGAATTATTGGGCGTAAAGCGCGTGTAGGCGGCGTG
ACAAGTCGGGTGTGAAAGCCCTCAGCTCAACTGAGGAAGTGCGCCCGAAACTGT
CGTGCTTGAGTGCCGGAGAGGGTGGCGGAATTCCCCAAGTAGAGGTGAAATTCG
TAGATATGGGGAGGAACACCGGTGGCGAAGGCGGCCACCTGGACGGTAACTGAC
GCTGAGACGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCAC
GCCGTAAACGATGAGAACTAGGTGTCGTGGGAGTTGACCCCCGCGGTGCCGTAG
CTAACGCATTAAGTTCTCCGCCTGGGAAGTACGGTCGCAAGACTAAAACTCAAAG
GAATTGACGGGGG

34. Contig of Mx2132KM

TAGTTCTCATCGTTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCC
ACGCTTTCGCGTCTCAGCGTCAGTTACCGTCCAGGTGGCCGCCTTCGCCACCGGT
GTTCCTCCCCATATCTACGAATTTCACCTCTACTTGGGGAATTCCGCCACCCTCTC
CGGCACTCAAGCACGACAGTTTCGGGCGCACTTCCTCAGTTGAGCTGAGGGCTTT
CACACCCGACTTGTCACGCCGCCTACACGCGCTTTACGCCCAATAATTCCGAACA
ACGCTTGCACCCTCTGTATTACCGCGGCTGCTGGCACAGAGTTAGCCGGTGCTTCT
TCTCCCGGTACCGTCAAGCCGTAGCGTGTTGGGCTACGGGTTTTCGTCCCGGTCG
AAAGTGCTTTACAATCCGAAGACCTTCATCACACACGCGGCGTTGCTGCGTCAGG
CTTTCGCCCATTGCGCAAAATTCCCCACTGCTGCCTCCCGTAGGAGTCTGGACCGT
GTCTCAGTTCCAGTGTGGCTGATCGTCCTCTCAGACCAGCTACCCGTCGTTGCCTT
GGTGGGCCATTACCCCGCCAACTAGCTGATGGGCCGCGGACTCATCTGGTTGTGA
TAGCTTGTATACAGAGGCCACCTTTTCCCTCAGTCTCCGAAGAAACCGTGGGCTT
ATCCGGTATTAGCCAATCTTTCGACTGGTTATCCCAGGCATCCAGGCAGATTATCC
ACGTGTTACGCACCCGTGCGCCGCTCTACTAAGGTTTCCCTATTCGCGCTCGACTT
GCATGTGTTAGGCACGCCGCCAGCGTTCGTTCTGAG

35. Contig of MxGP4KM, 846bp

TCAACTCCCACGACACCTAGTTCTCATCGTTTACGGCGTGGACTACCAGGGTATCT
AATCCTGTTTGCTCCCCACGCTTTCGCGTCTCAGCGTCAGTTACCGTCCAGGTGGC
CGCCTTCGCCACCGGTGTTCCTCCCCATATCTACGAATTTCACCTCTACTTGGGGA
ATTCCGCCACCCTCTCCGGCACTCAAGCACGACAGTTTCGGGCGCACTTCCTCAG
TTGAGCTGAGGGCTTTCACACCCGACTTGTCACGCCGCCTACACGCGCTTTACGC
CCAATAATTCCGAACAACGCTTGCACCCTCTGTATTACCGCGGCTGCTGGCACAG
AGTTAGCCGGTGCTTCTTCTCCCGGTACCGTCAAGCCGGAGCGTGTTAGGCTCCG
GGTTTTCGTCCCGGTCGAAAGTGCTTTACAATCCGAAGACCTTCATCACACACGC
GGCGTTGCTGCGTCAGGCTTTCGCCCATTGCGCAAAATTCCCCACTGCTGCCTCCC
GTAGGAGTCTGGACCGTGTCTCAGTTCCAGTGTGGCTGATCGTCCTCTCAGACCA
GCTACCCGTCGTCGCCTTGGTGGGCCATTACCCCGCCAACTAGCTGATGGGCCGC
GGACTCATCTGGTTGTGATAGCTTGTATACAGAGGCCACCTTTTCCCTCAGTCTCC
GAAGAAACCGTGGGCTTATCCGGTATTAGCCAATCTTTCGACTGGTTATCCCAGG
CATCCAGGCAGATTATCCACGTGTTACGCACCCGTGCGCCGCTCTACTAAGGGTT
CCCTATTCGCGCTCGACTTGCATGTGTTAGGCACGCCGCCAGCGTTCGTTCTGAGC
CATGATCAAACTCTA

36. Contig of CeGP43TM, 819bp

TAGTTCTCATCGTTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCC
ACGCTTTCGCGTCTCAGCGTCAGTTACCGTCCAGGTGGCCGCCTTCGCCACCGGT
GTTCCTCCCCATATCTACGAATTTCACCTCTACTTGGGGAATTCCGCCACCCTCTC

CGGCACTCAAGCACGACAGTTTCGGGCGCACTTCCTCAGTTGAGCTGAGGGCTTT
CACACCCGACTTGTCACGCCGCCTACACGCGCTTTACGCCCAATAATTCCGAACA
ACGCTTGCACCCTCTGTATTACCGCGGCTGCTGGCACAGAGTTAGCCGGTGCTTCT
TCTCCCGGTACCGTCAAGCCGTAGCGTGTTGGGCTACGGGTTTTCGTCCCGGTCG
AAAGTGCTTTACAATCCGAAGACCTTCATCACACACGCGGCGTTGCTGCGTCAGG
CTTTCGCCCATTGCGCAAAATTCCCCACTGCTGCCTCCCGTAGGAGTCTGGACCGT
GTCTCAGTTCCAGTGTGGCTGATCGTCCTCTCAGACCAGCTACCCGTCGTTGCCTT
GGTGGGCCATTACCCCGCCAACTAGCTGATGGGCCGCGGACTCATCTGGTTGTGA
TAGCTTGTATACAGAGGCCACCTTTTCCCTCAGTCTCCGAAGAAACCGTGGGCTT
ATCCGGTATTAGCCAATCTTTCGACTGGTTATCCCAGGCATCCAGGCAGATTATCC
ACGTGTTACGCACCCGTGCGCCGCTCTACTAAGGGTTGCCCCTATTCGCGCTCGA
CTTGCATGTGTTAGGCACGCCGCCAGCGTTCGTTCTGAGCCAT

37. Contig of MxMK7KM, 830bp

TCCCACGACACCTAGTTCTCATCGTTTACGGCGTGGACTACCAGGGTATCTAATCC
TGTTTGCTCCCCACGCTTTCGCGTCTCAGCGTCAGTTACCGTCCAGGTGGCCGCCT
TCGCCACCGGTGTTCCTCCCCATATCTACGAATTTCACCTCTACTTGGGGAATTCC
GCCACCCTCTCCGGCACTCAAGCACAACAGTTTCGGGCGCACTTCCTCAGTTGAG
CTGAGGGCTTTCACACCCGACTTGTCACGCCGCCTACACGCGCTTTACGCCCAAT
AATTCCGAACAACGCTTGCACCCTCTGTATTACCGCGGCTGCTGGCACAGAGTTA
GCCGGTGCTTCTTCTCCCGGTACCGTCAAGCCGTTGGATGTTAGCCAACGGGTTTT
CTTCCCGGTCGAAAGTGCTTTACAATCCAAAGACCTTCATCACACACGCGGCGTT
GCTGCGTCAGGCTTTCGCCCATTGCGCAAAATTCCCCACTGCTGCCTCCCGTAGG
AGTCTGGACCGTGTCTCAGTTCCAGTGTGGCTGATCGTCCTCTCAGACCAGCTACC
CGTCGTTGCCTTGGTGGGCCATTACCCCGCCAACTAGCTGATGGGCCGCGGACTC
ATCTGAATGTGATAGCTTGTATACAGAGGCCACCTTTTCCCTCAGTCTCCGAAGA
AACCGTGGGCTTATCCGGTATTAGCCAATCTTTCGACTGGTTATCCCGAGCACTCA
GGCAGATTATCCACGTGTTACGCACCCGTGCGCCGCTCTACTAAGGGTTGCCCCT
ATTCGCGCTCGACTTGCATGTGTTAGGCACGCCGCCAGCGTTCGTTCTGAGCCA

38. Contig of MxMK5-1KM, 426bp

GCTTTACAATCCAAAGACCTTCATCACACACGCGGCGTTGCTGCGTCAGGCTTTC
GCCCATTGCGCAAAATTCCCCACTGCTGCCTCCCGTAGGAGTCTGGACCGTGTCT
CAGTTCCAGTGTGGCTGATCGTCCTCTCAGACCAGCTACCCGTCGTTGCCTTGGTG
GGCCATTACCCCGCCAACTAGCTGATGGGCCGCGGACTCATCTGAATGTGATAGC
TTGTATACAGAGGCCACCTTTTCCCTCAGTCTCCGAAGAAACCGTGGGCTTATCC
GGTATTAGCCAATCTTTCGACTGGTTATCCCGAGCACTCAGGCAGATTATCCACG
TGTTACGCACCCGTGCGCCGCTCTACTAAGGGTTGCCCCTATTCGCGCTCGACTTG
CATGTGTTAGGCACGCCGCCAGCGTTCGTTCTGAGCCAT

39. Contig of MxGC9KM, 795bp

TCCCACGACACCTAGTTCTCATCGTTTACGGCGTGGACTACCAGGGTATCTAATCC
TGTTTGCTCCCCACGCTTTCGCGTCTCAGCGTCAGTTACCGTCCAGGTGGCCGCCT
TCGCCACCGGTGTTCCTCCCCATATCTACGAATTTCACCTCTACTTGGGGAATTCC
GCCACCCTCTCCGGCACTCAAGCACGACAGTTTCGGGCGCACTTCCTCAGTTGAG
CTGAGGGCTTTCACACCCGACTTGTCACGCCGCCTACACGCGCTTTACGCCCAAT
AATTCCGAACAACGCTTGCACCCTCTGTATTACCGCGGCTGCTGGCACAGAGTTA
GCCGGTGCTTCTTCTCCCGGTACCGTCAAGCCGTTGGATGTTAGCCAACGGGTTTT

CTTCCCGGTCGAAAGTGCTTTACAATCCAAAGACCTTCATCACACACGCGGCGTT
GCTGCGTCAGGCTTTCGCCCATTGCGCAAAATTCCCCACTGCTGCCTCCCGTAGG
AGTCTGGACCGTGTCTCAGTTCCAGTGTGGCTGATCGTCCTCTCAGACCAGCTACC
CGTCGTTGCCTTGGTGGGCCATTACCCCGCCAACTAGCTGATGGGCCGCGGACTC
ATCTGAATGTGATAGCTTGTATACAGAGGCCACCTTTTCCCTCAGTCTCCGAAGA
AACCGTGGGCTTATCCGGTATTAGCCAATCTTTCGACTGGTTATCCCGAGCACTCA
GGCAGATTATCCACGTGTTACGCACCCGTGCGCCGCTCTACTAAGGGTTGCCCCT
ATTCGCGCTCGACTTGCATGTGTTAGGCACGCCGCCAGCGTTCGTTCTGAGCCAT

40. Contig of Mx1941KM, 838bp

TCAACTCCCACGACACCTAGTTCTCATCGTTTACGGCGTGGACTACCAGGGTATCT
AATCCTGTTTGCTCCCCACGCTTTCGCGTCTCAGCGTCAGTTACCGTCCAGGTGGC
CGCCTTCGCCACCGGTGTTCCTCCCCATATCTACGAATTTCACCTCTACTTGGGGA
ATTCCGCCACCCTCTCCGGCACTCAAGCACGACAGTTTCGGGCGCACTTCCTCAG
TTGAGCTGAGGGCTTTCACACCCGACTTGTCACGCCGCCTACACGCGCTTTACGC
CCAATAATTCCGAACAACGCTTGCACCCTCTGTATTACCGCGGCTGCTGGCACAG
AGTTAGCCGGTGCTTCTTCTCCCGGTACCGTCAAGCCGTTGGATGTTAGCCAACG
GGTTTTCTTCCCGGTCGAAAGTGCTTTACAATCCAAAGACCTTCATCACACACGC
GGCGTTGCTGCGTCAGGCTTTCGCCCATTGCGCAAAATTCCCCACTGCTGCCTCCC
GTAGGAGTCTGGACCGTGTCTCAGTTCCAGTGTGGCTGATCGTCCTCTCAGACCA
GCTACCCGTCGTTGCCTTGGTGGGCCATTACCCCGCCAACTAGCTGATGGGCCGC
GGACTCATCTGAATGTGATAGCTTGTATACAGAGGCCACCTTTTCCCTCAGTCTCC
GAAGAAACCGTGGGCTTATCCGGTATTAGCCAATCTTTCGACTGGTTATCCCGAG
CACTCAGGCAGATTATCCACGTGTTACGCACCCGTGCGCCGCTCTACTAAGGGTT
GCCCCTATTCGCGCTCGACTTGCATGTGTTAGGCACGCCGCCAGCGTTCGTTCTGA
GCCATGA

41. Contig of MxGP1962KM, 835bp

TCAACTCCCACGACACCTAGTTCTCATCGTTTACGGCGTGGACTACCAGGGTATCT
AATCCTGTTTGCTCCCCACGCTTTCGCGTCTCAGCGTCAGTTACCGTCCAGGTGGC
CGCCTTCGCCACCGGTGTTCCTCCCCATATCTACGAATTTCACCTCTACTTGGGGA
ATTCCGCCACCCTCTCCGGCACTCAAGCACGACAGTTTCGGGCGCACTTCCTCAG
TTGAGCTGAGGGCTTTCACACCCGACTTGTCACGCCGCCTACACGCGCTTTACGC
CCAATAATTCCGAACAACGCTTGCACCCTCTGTATTACCGCGGCTGCTGGCACAG
AGTTAGCCGGTGCTTCTTCTCCCGGTACCGTCAAGCTGTTGGATGTTAGCCAACA
GGGTTTCTTCCCGGTCGAAAGTGCTTTACAATCCAAAGACCTTCATCACACACGC
GGCGTTGCTGCGTCAGGCTTTCGCCCATTGCGCAAAATTCCCCACTGCTGCCTCCC
GTAGGAGTCTGGACCGTGTCTCAGTTCCAGTGTGGCTGATCGTCCTCTCAGACCA
GCTACCCGTCGTTGCCTTGGTGGGCCATTACCCCGCCAACTAGCTGATGGGCCGC
GGACTCATCTGAATGTGATAGCTTGTATACAGAGGCCACCTTTTCCCTCAGTCTCC
GAAGAAACCGTGGGCTTATCCGGTATTAGCCAATCTTTCGACTGGTTATCCCGAG
CACTCAGGCAGATTATCCACGTGTTACGCACCCGTGCGCCGCTCTACTAAGGGTT
GCCCCTATTCGCGCTCGACTTGCATGTGTTAGGCACGCCGCCAGCGTTCGTTCTGA
GCCA

42. Contig of Mxx1929(1) KR, 934bp

TCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTTAGTCTTGCGACCGTA
CTTCCCAGGCGGAGAACTTAATGCGTTAGCTTCGGCACCGCGGGGGTCAACTCCC
ACGACACCTAGTTCTCATCGTTTACGGCGTGGACTACCAGGGTATCTAATCCTGTT
TGCTCCCCACGCTTTCGCGTCTCAGCGTCAGTTACCGTCCAGGTGGCCGCCTTCGC
CACCGGTGTTCCTCCCCATATCTACGAATTTCACCTCTACTTGGGGAATTCCGCCA
CCCTCTCCGGCACTCAAGCACGACAGTTTCGGGCGCACTTCCTCAGTTGAGCTGA
GGGCTTTCACACCCGACTTGTCACGCCGCCTACACGCGCTTTACGCCCAATAATTC
CGAACAACGCTTGCACCCTCTGTATTACCGCGGCTGCTGGCACAGAGTTAGCCGG
TGCTTCTTCTCCCGGTACCGTCAAGCTGTTGGATGTTAGCCAACAGGGTTTCTTCC
CGGTCGAAAGTGCTTTACAATCCAAAGACCTTCATCACACACGCGGCGTTGCTGC
GTCAGGCTTTCGCCCATTGCGCAAAATTCCCCACTGCTGCCTCCCGTAGGAGTCTG
GACCGTGTCTCAGTTCCAGTGTGGCTGATCGTCCTCTCAGACCAGCTACCCGTCGT
TGCCTTGGTGGGCCATTACCCCGCCAACTAGCTGATGGGCCGCGGACTCATCTGA
ATGTGATAGCTTGTATACAGAGGCCACCTTTTCCCTCAGTCTCCGAAGAAACCGT
GGGCTTATCCGGTATTAGCCAATCTTTCGACTGGTTATCCCGAGCACTCAGGCAG
ATTATCCACGTGTTACGCACCCGTGCGCCGCTCTACTAAGGGTTGCCCCTATTCGC
GCTCGACTTGCATGTGTTAGGCACGCCGCCAGCGTTCGTTCTGAG

43. Contig of Mxx1929KR, 835bp

TCAACTCCCACGACACCTAGTTCTCATCGTTTACGGCGTGGACTACCAGGGTATCT
AATCCTGTTTGCTCCCCACGCTTTCGCGTCTCAGCGTCAGTTACCGTCCAGGTGGC
CGCCTTCGCCACCGGTGTTCCTCCCCATATCTACGAATTTCACCTCTACTTGGGGA
ATTCCGCCACCCTCTCCGGCACTCAAGCACGACAGTTTCGGGCGCACTTCCTCAG
TTGAGCTGAGGGCTTTCACACCCGACTTGTCACGCCGCCTACACGCGCTTTACGC
CCAATAATTCCGAACAACGCTTGCACCCTCTGTATTACCGCGGCTGCTGGCACAG
AGTTAGCCGGTGCTTCTTCTCCCNGTACCGTCAAGCTGTTGGATGTTAGCCAACA
GGGTTTCTTCCCGGTCGAAAGTGCTTTACAATCCAAAGACCTTCATCACACACGC
GGCGTTGCTGCGTCAGGCTTTCGCCCATTGCGCAAAATTCCCCACTGCTGCCTCCC
GTAGGAGTCTGGACCGTGTCTCAGTTCCAGTGTGGCTGATCGTCCTCTCAGACCA
GCTACCCGTCGTTGCCTTGGTGGGCCATTACCCCGCCAACTAGCTGATGGGCCGC
GGACTCATCTGAATGTGATAGCTTGTATACAGAGGCCACCTTTTCCCTCAGTCTCC
GAAGAAACCGTGGGCTTATCCGGTATTAGCCAATCTTTCGACTGGTTATCCCGAG
CACTCAGGCAGATTATCCACGTGTTACGCACCCGTGCGCCGCTCTACTAAGGGTT
GCCCCTATTCGCGCTCGACTTGCATGTGTTAGGCACGCCGCCAGCGTTCGTTCTGA
GCCA

44. Contig of Mx1955KM, 834bp

TCAACTCCCACGACACCTAGTTCTCATCGTTTACGGCGTGGACTACCAGGGTATCT
AATCCTGTTTGCTCCCCACGCTTTCGCGTCTCAGCGTCAGTTACCGTCCAGGTGGC
CGCCTTCGCCACCGGTGTTCCTCCCCATATCTACGAATTTCACCTCTACTTGGGGA
ATTCCGCCACCCTCTCCGGCACTCAAGCACGACAGTTTCGGGCGCACTTCCTCAG
TTGAGCTGAGGGCTTTCACACCCGACTTGTCACGCCGCCTACACGCGCTTTACGC
CCAATAATTCCGAACAACGCTTGCACCCTCTGTATTACCGCGGCTGCTGGCACAG
AGTTAGCCGGTGCTTCTTCTCCCGGTACCGTCAAGCTGTTGGATGTTAGCCAACA
GGGTTTCTTCCCGGTCGAAAGTGCTTTACAATCCAAAGACCTTCATCACACACGC
GGCGTTGCTGCGTCAGGCTTTCGCCCATTGCGCAAAATTCCCCACTGCTGCCTCCC
GTAGGAGTCTGGACCGTGTCTCAGTTCCAGTGTGGCTGATCGTCCTCTCAGACCA

GCTACCCGTCGTTGCCTTGGTGGGCCATTACCCCGCCAACTAGCTGATGGGCCGC
GGACTCATCTGAATGTGATAGCTTGTATACAGAGGCCACCTTTTCCCTCAGTCTCC
GAAGAAACCGTGGGCTTATCCGGTATTAGCCAATCTTTCGACTGGTTATCCCGAG
CACTCAGGCAGATTATCCACGTGTTACGCACCCGTGCGCCGCTCTACTAAGGGTT
GCCCCTATTCGCGCTCGACTTGCATGTGTTAGGCACGCCGCCAGCGTTCGTTCTGA
GCC

45. Contig of TP13RAWA-A, 939bp

TCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTTAGTCTTGCGACCGTA
CTTCCCAGGCGGAGAACTTAATGCGTTAGCTTCGGCACCGCGGGGGTCAACTCCC
ACGACACCTAGTTCTCATCGTTTACGGCGTGGACTACCAGGGTATCTAATCCTGTT
TGCTCCCCACGCTTTCGCGTCTCAGCGTCAGTTACCGTCCAGGTGGCCGCCTTCGC
CACCGGTGTTCCTCCCCATATCTACGAATTTCACCTCTACTTGGGGAATTCCGCCA
CCCTCTCCGGCACTCAAGCACAACAGTTTCGGGCGCACTTCCTCAGTTGAGCTGA
GGGCTTTCACACCCGACTTGTCACGCCGCCTACACGCGCTTTACGCCCAATAATTC
CGAACAACGCTTGCACCCTCTGTATTACCGCGGCTGCTGGCACAGAGTTAGCCGG
TGCTTCTTCTCCCGGTACCGTCAAGCCGTTGGATGTTAGCCAACGGGTTTTCTTCC
CGGTCGAAAGTGCTTTACAATCCAAAGACCTTCATCACACACGCGGCGTTGCTGC
GTCAGGCTTTCGCCCATTGCGCAAAATTCCCCACTGCTGCCTCCCGTAGGAGTCTG
GACCGTGTCTCAGTTCCAGTGTGGCTGATCGTCCTCTCAGACCAGCTACCCGTCGT
TGCCTTGGTGGGCCATTACCCCGCCAACTAGCTGATGGGCCGCGGACTCATCTGA
ATGTGATAGCTTGTATACAGAGGCCACCTTTTCCCTCAGTCTCCGAAGAAACCGT
GGGCTTATCCGGTATTAGCCAATCTTTCGACTGGTTATCCCGAGCACTCAGGCAG
ATTATCCACGTGTTACGCACCCGTGCGCCGCTCTACTAAGGGTTGCCCCTATTCGC
GCTCGACTTGCATGTGTTAGGCACGCCGCCAGCGTTCGTTCTGAGCCATG

46. Contig of TP13RAWA-B, 817bp

TAGTTCTCATCGTTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCC
ACGCTTTCGCGTCTCAGCGTCAGTTACCGTCCAGGTGGCCGCCTTCGCCACCGGT
GTTCCTCCCCATATCTACGAATTTCACCTCTACTTGGGGAATTCCGCCACCCTCTC
CGGCACTCAAGCACAACAGTTTCGGGCGCACTTCCTCAGTTGAGCTGAGGGCTTT
CACACCCGACTTGTCACGCCGCCTACACGCGCTTTACGCCCAATAATTCCGAACA
ACGCTTGCACCCTCTGTATTACCGCGGCTGCTGGCACAGAGTTAGCCGGTGCTTCT
TCTCCCGGTACCGTCAAGCCGTTGGATGTTAGCCAACGGGTTTTCTTCCCGGTCGA
AAGTGCTTTACAATCCAAAGACCTTCATCACACACGCGGCGTTGCTGCGTCAGGC
TTTCGCCCATTGCGCAAAATTCCCCACTGCTGCCTCCCGTAGGAGTCTGGACCGTG
TCTCAGTTCCAGTGTGGCTGATCGTCCTCTCAGACCAGCTACCCGTCGTTGCCTTG
GTGGGCCATTACCCCGCCAACTAGCTGATGGGCCGCGGACTCATCTGAATGTGAT
AGCTTGTATACAGAGGCCACCTTTTCCCTCAGTCTCCGAAGAAACCGTGGGCTTA
TCCGGTATTAGCCAATCTTTCGACTGGTTATCCCGAGCACTCAGGCAGATTATCCA
CGTGTTACGCACCCGTGCGCCGCTCTACTAAGGGTTGCCCCTATTCGCGCTCGACT
TGCATGTGTTAGGCACGCCGCCAGCGTTCGTTCTGAGCC

47. Contig of TP13RAWA-C, 834bp

TCAACTCCCACGACACNTAGTTCTCATCGTTTACGGCGTGGACTACCAGGGTATC
TAATCCTGTTTGCTCCCCACGCTTTCGCGTCTCAGCGTCAGTTACCGTCCAGGTGG
CCGCCTTCGCCACCGGTGTTCCTCCCCATATCTACGAATTTCACCTCTACTTGGGG
AATTCCGCCACCCTCTCCGGCACTCAAGCACAACAGTTTCGGGCGCACTTCCTCA

GTTGAGCTGAGGGCTTTCACACCCGACTTGTCACGCCGCCTACACGCGCTTTACG
CCCAATAATTCCGAACAACGCTTGCACCCTCTGTATTACCGCGGCTGCTGGCACA
GAGTTAGCCGGTGCTTCTTCTCCCGGTACCGTCAAGCCGTTGGATGTTAGCCAAC
GGGTTTTCTTCCCGGTCGAAAGTGCTTTACAATCCAAAGACCTTCATCACACACG
CGGCGTTGCTGCGTCAGGCTTTCGCCCATTGCGCAAAATTCCCCACTGCTGCCTCC
CGTAGGAGTCTGGACCGTGTCTCAGTTCCAGTGTGGCTGATCGTCCTCTCAGACC
AGCTACCCGTCGTTGCCTTGGTGGGCCATTACCCCGCCAACTAGCTGATGGGCCG
CGGACTCATCTGAATGTGATAGCTTGTATACAGAGGCCACCTTTTCCCTCAGTCTC
CGAAGAAACCGTGGGCTTATCCGGTATTAGCCAATCTTTCGACTGGTTATCCCGA
GCACTCAGGCAGATTATCCACGTGTTACGCACCCGTGCGCCGCTCTACTAAGGGT
TGCCCCTATTCGCGCTCGACTTGCATGTGTTAGGCACGCCGCCAGCGTTCGTTCTG
AGCC

48. Contig of Mx1960KM, 750bp

TCAGCGTCAGTTACCGTCCAGGTGGCCGCCTTCGCCACCGGTGTTCCTCCCCATAT
CTACGAATTTCACCTCTACTTGGGGAATTCCGCCACCCTCTCCGGCACTCAAGCAC
GACAGTTTCGGGCGCACTTCCTCAGTTGAGCTGAGGGCTTTCACACCCGACTTGT
CACGCCGCCTACACGCGCTTTACGCCCAATAATTCCGAACAACGCTTGCACCCTC
TGTATTACCGCGGCTGCTGGCACAGAGTTAGCCGGTGCTTCTTCTCCCGGTACCGT
CAAGCTGTTGGATGTTAGCCAACAGGGTTTCTTCCCGGTCGAAAGTGCTTTACAA
TCCAAAGACCTTCATCACACACGCGGCGTTGCTGCGTCAGGCTTTCGCCCATTGC
GCAAAATTCCCCACTGCTGCCTCCCGTAGGAGTCTGGACCGTGTCTCAGTTCCAG
TGTGGCTGATCGTCCTCTCAGACCAGCTACCCGTCGTTGCCTTGGTGGGCCATTAC
CCCGCCAACTAGCTGATGGGCCGCGGACTCATCTGAATGTGATAGCTTGTATACA
GAGGCCACCTTTTCCCTCAGTCTCCGAAGAAACCGTGGGCTTATCCGGTATTAGC
CAATCTTTCGACTGGTTATCCCGAGCACTCAGGCAGATTATCCACGTGTTACGCA
CCCGTGCGCCGCTCTACTAAGGGTTGCCCCTATTCGCGCTCGACTTGCATGTGTTA
GGCACGCCGCCAGCGTTCGTTCTGAGCCAT

49. Contig of Mx1960KR(1), 835bp

TCAACTCCCACGACACCTAGTTCTCATCGTTTACGGCGTGGACTACCAGGGTATCT
AATCCTGTTTGCTCCCCACGCTTTCGCGTCTCAGCGTCAGTTACCGTCCAGGTGGC
CGCCTTCGCCACCGGTGTTCCTCCCCATATCTACGAATTTCACCTCTACTTGGGGA
ATTCCGCCACCCTCTCCGGCACTCAAGCACGACAGTTTCGGGCGCACTTCCTCAG
TTGAGCTGAGGGCTTTCACACCCGACTTGTCACGCCGCCTACACGCGCTTTACGC
CCAATAATTCCGAACAACGCTTGCACCCTCTGTATTACCGCGGCTGCTGGCACAG
AGTTAGCCGGTGCTTCTTCTCCCGGTACCGTCAAGCTGTTGGATGTTAGCCAACA
GGGTTTCTTCCCGGTCGAAAGTGCTTTACAATCCAAAGACCTTCATCACACACGC
GGCGTTGCTGCGTCAGGCTTTCGCCCATTGCGCAAAATTCCCCACTGCTGCCTCCC
GTAGGAGTCTGGACCGTGTCTCAGTTCCAGTGTGGCTGATCGTCCTCTCAGACCA
GCTACCCGTCGTTGCCTTGGTGGGCCATTACCCCGCCAACTAGCTGATGGGCCGC
GGACTCATCTGAATGTGATAGCTTGTATACAGAGGCCACCTTTTCCCTCAGTCTCC
GAAGAAACCGTGGGCTTATCCGGTATTAGCCAATCTTTCGACTGGTTATCCCGAG
CACTCAGGCAGATTATCCACGTGCTTACGCACCCGTGCGCCGCTCTACTAAGGGT
TGCCCCTATTCGCGCTCGACTTGCATGTGTTAGGCACGCCGCCAGCGTTCGTTCTG
AGCC

50. Contig of Mx2133KM, 891bp

51. Contig of MxGP1941KM, 929bp

52. Contig of Mx2129KM, 834bp

53. Contig of MxGP1963KM, 905bp

54. Contig of Mx1956KM, 895bp

55. Contig of Mx1965KR, 922bp

56. Contig of Mx1809KM, 897bp

57. Contig of soce1964KM, 886bp

58. Contig of 1932KM, 889bp

59. Contig of A1932RA bla1, 531bp

3.2 Full sequence 16S rRNA gene sequence

1. Contig of MxMK7KM, 1507bp

2. Contig of CcGP6KM, 1504bp

3. Contig of CCe1965KR, 1506bp

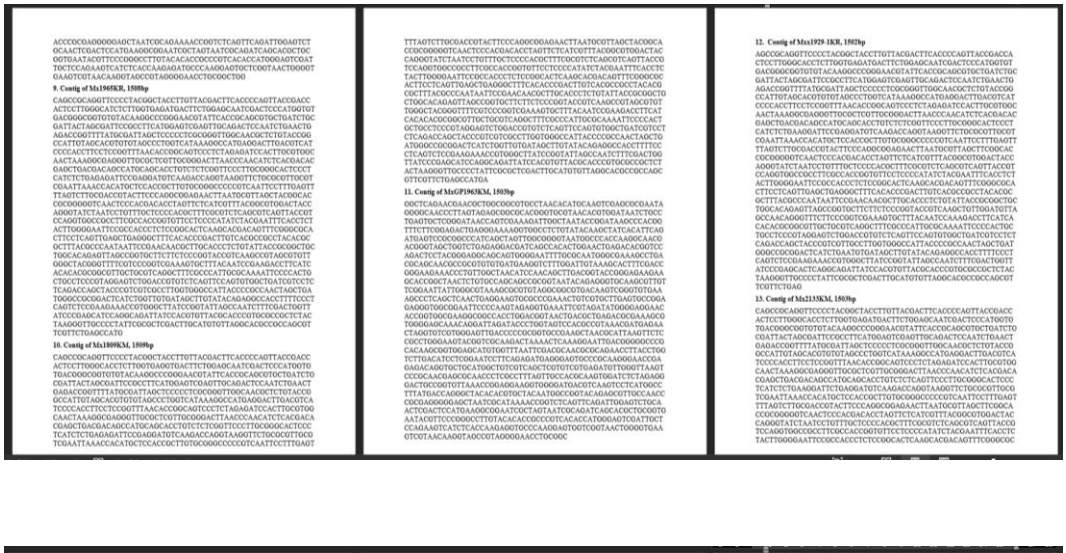
4. Contig of soce1969-2KR, 1506bp

5. Contig of CCMK6KM, 1506bp

6. MyxoGP11TM, 1452bp

7. Contig of TP13RAWA-A, 1515bp

8. Contig of 3RB1-1, 1510bp

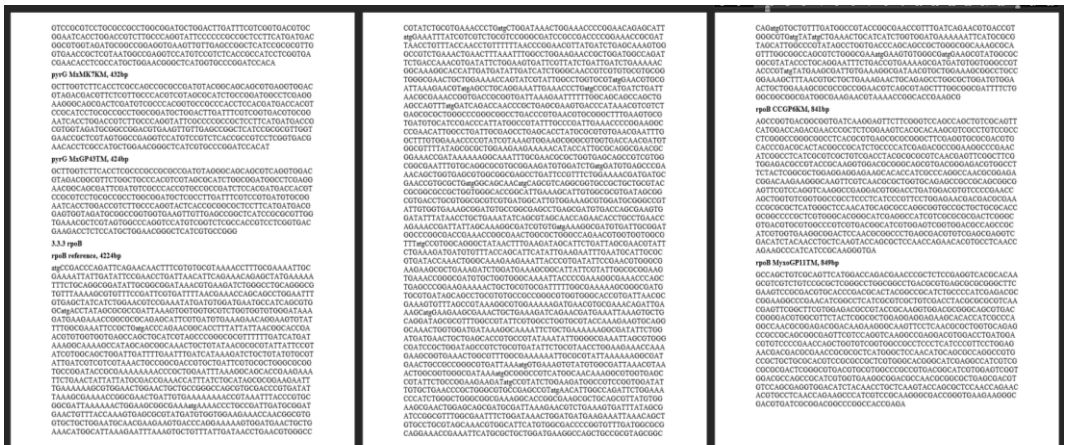


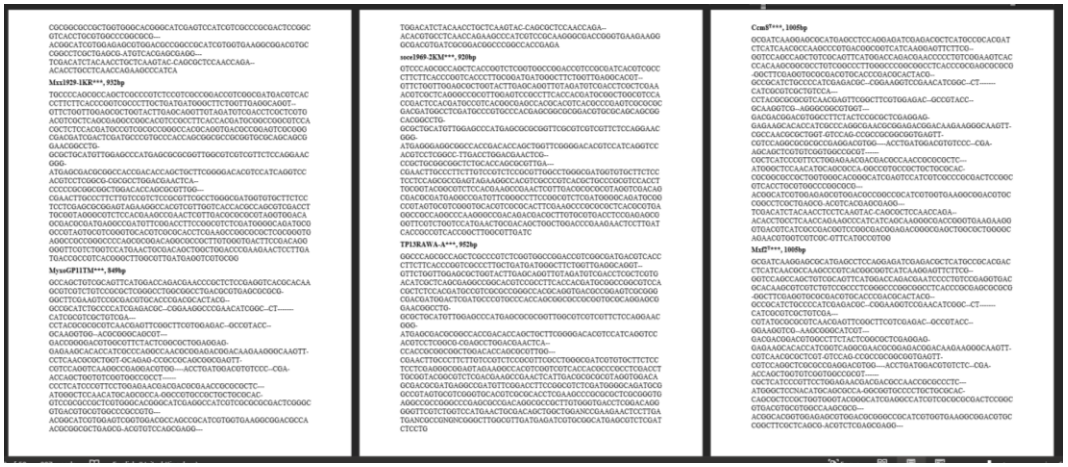
ACCCCCCAGGGGGAGCTAATCGCAGAAAACCGGTCTCAGTTCAGAATTGGAGTCT
GCAACTCGACTCCATGAAGCGGAATCGCTAGTAATCGCAGCATCAGCAGCACGC
GGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTCGAT
TGCTCCAGAAGTAGTTAACCGAAGGATCTAACCGCAAGGAGGCCGCTAACCTTCG
GGAGCGGTTTGTGTACAAGGCCCGGGAACGTATTCACCGCAGCGTTCTGATCGC
GAAGCTAACGCATTAAGTCGCCTCCGGGGTTGGCAGCCTCTGTCACCGGCTGAC
TGGCACAGAGTTAGCCGGTGCTTCTTCTGCGGGTAACGTCAATGAGCAAAGGTA
TTAGTCTTTGCAGCCATGACTTCCCCAGGCCGAAGCTTAAGTCCTCATCGCGTC
GCAAGACCAAAGAGGGGGACCTTCGGGCCTCTTGCCATCGGATGTGCCCAGATG
GGATTAGCTAGTAGGTGGGGTAACGGCTCACCTAGGCGACGATCCCTAGCTGGT
CTGAGAGGATGACCAGCCACACTGGAACTGAGACACGGTCCAGACTCCTACGGG
AGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCG
CGTGTATGAAGAAGGCCTTCGGGTTGTAAAGTACTTTCAGCGGGGAGGAAGGG
AGTAAAGTTAATACCTTTGCTCATTGACGTTACCCGCAGAAGAAGCACCGGCTA
ACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCGAGCGTTGTCCGGAATT
ATTGGGCGTAAAGAGCTCGTAGGCGGTTTGTCGCGTCTGCTGTGAAAGCCCGG
GGCTTAACCTCGGGCCTGCAGTGGGTACGGGCAGACTAGAGTGCAGTAGGGGA
GACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCG
ATGGCGAAGGCAGGTCTCTGGGCTGTAACTGACGCTGAGGAGCGAAAGCGTGG
GGAGCGAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTA
AGTGTTAGGGGGTTTCCGCCCCTTAGTGCTGCAGCTAACGCATTAAGCACTCCG
CCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGC
ACAAGCGGTGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACCTGG
TCTTGACATCCACAGAATCTTCCAGAGATGGATTGGTGCCTTCGGGAACTGTGA
GACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGT
CCCGCAACGAGCGCAACCCTTGTCCTTAGTTGCCAGCATTAAGTTGGGCACTCT
AAGGGTACCGCCGGTGATAAGCGGAGGAAGGTGGGGATGACGTCAAGTCATCA
TGGCCCTTACGGCCAGGGCTACACACGTGCTACAATGGTCGGTACAGAGGGTTG
CCAAGCCGCGAGGTGGAGCTAATCCCAGAAAAGCCGATCGTAGTCCGGATCGC
AGTCTGCAACTCGACTGCGTGAAGTCGGAATCGCTAGTAATCGCGAATCAGAAT
GTCGCGGTGAATACGTTCCCGGGTCTTGTACACACCGCCCGTCACACCATGGGA
GTTGGTTGCACCAGAAGTAGCTAGTCTAACCTTCGGGAGGACGGTTACCACGG
TGTGATTCATGACTGGGGTGAAGTCGTAACAAGGTAACCGTAGGGGAACCTGC
GGTTGGATCACCTCCTTA

ACTTCCTCAGTTGAGCTGAGGGCTTTCACACCCGACTTGTCACGCCGCCTACCAG
CGCTTACGCCCAATAATTCCGAACAACGCTTGCACCCTCTGTATTACCGCGGCTG
CTGGCACAGAGTTAGCCGGTGCTTCTTCTGCGGGTAACGTCAATGAGCAAAGGT
ATTAGCTTGTTGCAGCCATGACTTCCCAGGCCGAAGCTTAAGTCCTCATCGCGT
CGCAAGACCAAAGAGGGGGACCTTCGGGCCTCTTGCCATCGGATGTGCCCAGA
TGGGATTAGCTAGTAGGTGGGGTAACGGCTCACCTAGGCGACGATCCCTAGCTG
GTCTGAGAGGATGACCAGCCACACTGGAACTGAGACACGGTCCAGACTCCTACG
GGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATG
CCGCGTGTATGAAGAAGGCCTTCGGGTTGTAAAGTACTTTCAGCGGGGAGGAA
GGGAGTAAAGTTAATACCTTTGCTCATTGACGTTACCCGCAGAAGAAGCACCGG
CTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCGAGCGTTGTCCGGA
ATTATTGGGCGTAAAGAGCTCGTAGGCGGTTTGTCGCGTCTGCTGTGAAAGCC
CGGGGCTTAACCTCGGGCCTGCAGTGGGTACGGGCAGACTAGAGTGCAGTAGG
GGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACA
CCGATGGCGAAGGCAGGTCTCTGGGCTGTAACTGACGCTGAGGAGCGAAAGCG
TGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTG
CTAAGTGTTAGGGGGTTTCCGCCCCTTAGTGCTGCAGCTAACGCATTAAGCACT
CCGCCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCC
CGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACC
TGGTCTTGACATCCACAGAATCTTCCAGAGATGGATTGGTGCCTTCGGGAACTG
TGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTA
AGTCCCGCAACGAGCGCAACCCTTGTCCTTAGTTGCCAGCATTAAGTTGGGCA
CTCTAAGGGTACCGCCGGTGATAAGCGGAGGAAGGTGGGGATGACGTCAAGTC
ATCATGGCCCTTACGGCCAGGGCTACACACGTGCTACAATGGTCGGTACAGAGG
GTTGCCAAGCCGCGAGGTGGAGCTAATCCCAGAAAAGCCGATCGTAGTCCGGA
TCGCAGTCTGCAACTCGACTGCGTGAAGTCGGAATCGCTAGTAATCGCGAATCA
GAATGTCGCGGTGAATACGTTCCCGGGTCTTGTACACACCGCCCGTCACACCAT
GGGAGTTGGTTGCACCAGAAGTAGCTAGTCTAACCTTCGGGAGGACGGTTACC
ACGGTGTGATTCATGACTGGGGTGAAGTCGTAACAAGGTAACCGTAGGGGAAC
CTGCGGTTGGATCACCTCCTTA

ATGGCGCATACAGAGGGGCAGCTAGCTGGCAACAGGATGCGAATCTCAAAAAGTG
CGTCTCAGTTCGGATTGGGGTCTGCAACTCGACACATGAAGTTGGAATCGCTAGT
TAATCGCGCATCAGCAATGGCGCGGTGAATACGTTCCCGGACTTGTACACACCG
CCCGTCAAGCCATGGGAGTTGGGTGGACCTGAAGATTGTGACCATTTAGGAGCAG
TTTAGGGTCTAACCACGCGCTGGGGATGAGCGTCGTCATGGCATTGTGACCCG
GCTCCGTCTGCAGGGTTTCTTCTCCGGTATTATCTGGATTATCTATCTGT
ATCGTACACCGTCCTGGCTGGAAACCGGGCACCGTGGTGCCTACCTGACGCTGA
GGAGGAGCTCCCAGCCAGCAGTCGGTAGAGGATCTGACGGGTACTCCGCACCAA
GTCGGTAGTGTGGATTGACGTGCGAAAGGGCGATGATCAGCCATTCACGTCGTC
TTGCGAAGCCAAGTCCTTTCTGGTCGGTCTTGCGATGATTATCATATAATTCTG
TATGAAAGGTACCGAATACCAAGCCTGTGCGAGTACACACCGCCCGTCACACCA
TGAGAGTTTGTAACACCCAAAGTCGGTGGGCCTAACCTTCGGGAGGACGGTGC
TGTGGTGGGCTAATGACTGGGGTGAAGTCGTAACAAGGTAGCCGTACTGGAAG
GTGCGGCTGGATCACCTCCTTT

GTAGCAGCTTGACGGTACGTATGAATAAGCACCGGCTAACTCCGTGCCAGCAGC
CGCGGTAATACAGAGGGTGCAAGCGTTGTCCGGAATTATTGGGTTTAAAGGGTC
GTAGGCGGCTTATTAAGTCAGTGGTGAAAGTTTGCAGCTTAACTGTAAAAGTGC
ATTGATACTGGTAAGCTTGAGTACTGTAGAGGTGGGCGGAATTCCGCATGTAGCG
GTGAAATGCATAGATATGCGGAGGAACACCGATGGCGAAGGCAGATCTACTGGA
AGTAAACTGACGCTGAGGTACGAAAGCGTGGGGATCGAACAGGATTAGATACCCT
GGTAGTCCACGCTGTAAACGATGAGTGCTAAGTGTTAGGGGGTTTCCGCCCCTT
AGTGCTGCAGCTAACGCATTAAGCACTCCGCCTGGGGAGTACGGCCGCAAGGTT
GAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAA
TTCGATGCAACGCGAAGAACCTTACCTGGTCTTGACATCCACAGAATCTTCCAGA
GATGGATTGGTGCCTTCGGGAACTGTGAGACAGGTGCTGCATGGCTGTCGTCAG
CTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTCCTT
AGTTGCCAGCATTAAGTTGGGCACTCTAAGGGTACCGCCGGTGATAAGCGGAGG
AAGGTGGGGATGACGTCAAGTCATCATGGCCCTTACGGCCAGGGCTACACACGT
GCTACAATGGTCGGTACAGAGGGTTGCCAAGCCGCGAGGTGGAGCTAATCCCA
GAAAGCCGATCGTAGTCCGGATCGCAGTCTGCAACTCGACTGCGTGAAGTCGGA
ATCGCTAGTAATCGCGAATCAGAATGTCGCGGTGAATACGTTCCCGGGTCTTGT
ACACACCGCCCGTCACACCATGGGAGTTGGTTGCACCAGAAGTAGCTAGTCTAA
CCTTCGGGAGGACGGTTACCACGGTGTGATTCATGACTGGGGTGAAGTCGTAAC
AAGGTAACCGTAGGGGAACCTGCGGTTGGATCACCTCCTTT





Supplement 5. Result of the Identity matrix, conserved regions and single nucleotides analysis for Chapter 3

A. Sequence 1: 1932KM, 1488

Sequence 2: DSM 103551^{T}, 1446

A. 1 Sequence Identity Matrix

BioEdit version 7.2.5 (11/12/2013)

Input Alignment File: F:\balibacter\HKG analysis for 1932KM\HKG_1932KM_dsm103551^{T}_multiple alignment.fas

Seq->	HKG	1932KM	DSM 103551^{T}
HKG	ID	0.129	0.159
1932KM	0.129	ID	0.303
DSM 103551^{T}	0.159	0.303	ID

A.2 Conserved Regions

BioEdit version 7.2.5 (11/12/2013)

Conserved regions search.

Alignment file: F:\balibacter\HKG analysis for 1932KM\HKG_1932KM_dsm103551^{T}_multiple alignment.fas

Minimum segment length (actual for each sequence): 3

Maximum average entropy: 0.2

Maximum entropy per position: 0.2

Gaps limited to 2 per segment.

Contiguous gaps limited to 1 in any segment.

4 conserved regions found.

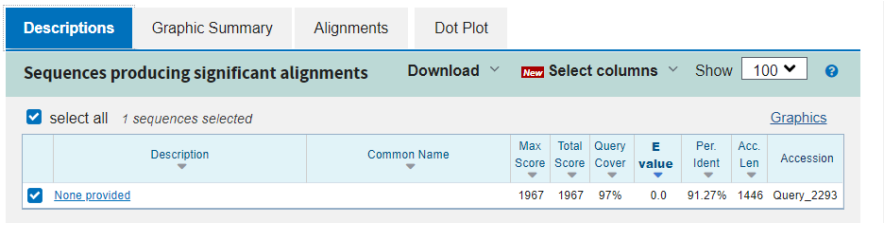
Region 1: 3326 AGT 3328 Segment Length: 3

Region 2: 3359 GTG 3361 Segment Length: 3

Region 3: 3799 CAGA 3802 Segment Length: 4

Region 4: 3876 AAC 3878 Segment Length: 3

A.3 Single Nucleotides Analysis Base On BLAST



Single nucleotides analysis: 109 positions

B. Sequence 1: 1932KM

Sequence 2: DSM 6795^{T}

B.1 Sequence Identity Matrix

BioEdit version 7.2.5 (11/12/2013)

Input Alignment File: F:\balibacter\HKG analysis for 1932KM\HKG_1932KM_dsm6795^{T}_multiple alignment.fas

Seq->	HKG	1932KM	DSM 6795^{T}
HKG	ID	0.134	0.164
1932KM	0.134	ID	0.333
DSM 6795^{T}	0.164	0.333	ID

B.2 Conserved Regions

BioEdit version 7.2.5 (11/12/2013)

Conserved regions search.

Alignment file: F:\balibacter\HKG analysis for 1932KM\HKG_1932KM_dsm6795^{T}_multiple alignment.fas

Minimum segment length (actual for each sequence): 3

Maximum average entropy: 0.2

Maximum entropy per position: 0.2

Gaps limited to 2 per segment.

Contiguous gaps limited to 1 in any segment.

7 conserved regions found.

Region 1: 2020 GCG 2022 Segment Length: 3

Region 2: 2115 TCC 2117 Segment Length: 3

Region 3: 2172 GCG 2174 Segment Length: 3

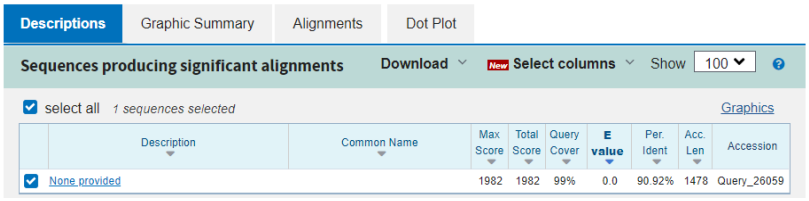
Region 4: 2344 CCT 2346 Segment Length: 3

Region 5: 2646 TGC 2648 Segment Length: 3

Region 6: 2813 AAC 2815 Segment Length: 3

Region 7: 2955 AAGG 2958 Segment Length: 4

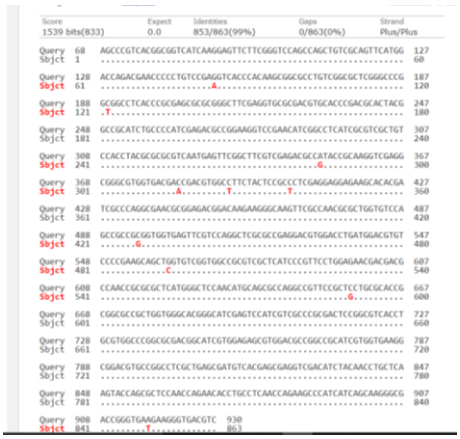
B.3 Single Nucleotides Analysis Base On BLAST



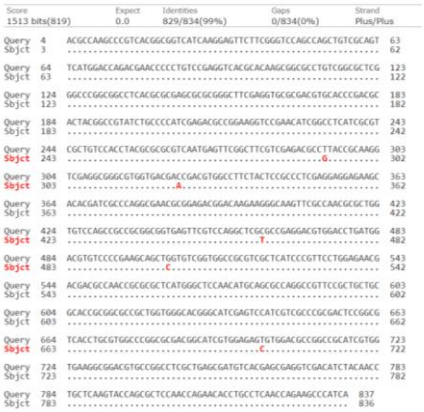
Single nucleotides analysis: 122 positions.

Supplement 6. Single nucleotide analysis for Chapter 4

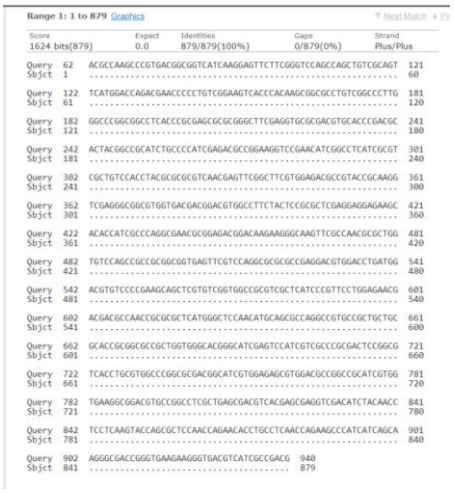
1. Sequence 1: Mxfl1***
   Sequence 2: sbmx085***



2. Sequence 1: dsm2260***
   Sequence 2: MxMK7KM***

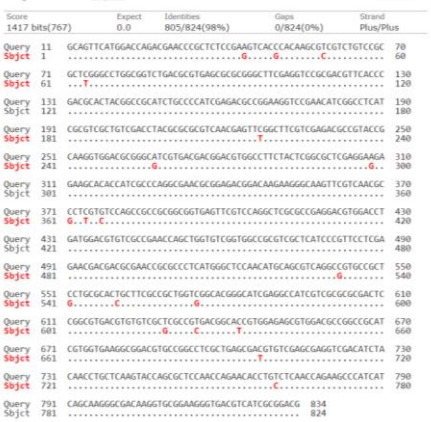


3. Sequence 1: Ccm8***
   Sequence 2: ccm8***



4. Sequence 1: MxGP43TM***
   Sequence 2: MxGP53TM***

5. Sequence 1: mxs8**
   Sequence 2: BGB2139RAST-03***



6. Sequence 1: sbco036
   Sequence 2: MyxoGP11TM

7. Sequence 1: mxf***
   Sequence 2: Mx1809KM***

8. Sequence 1: 3RB1-1***
   Sequence 2: CCe1965KR***

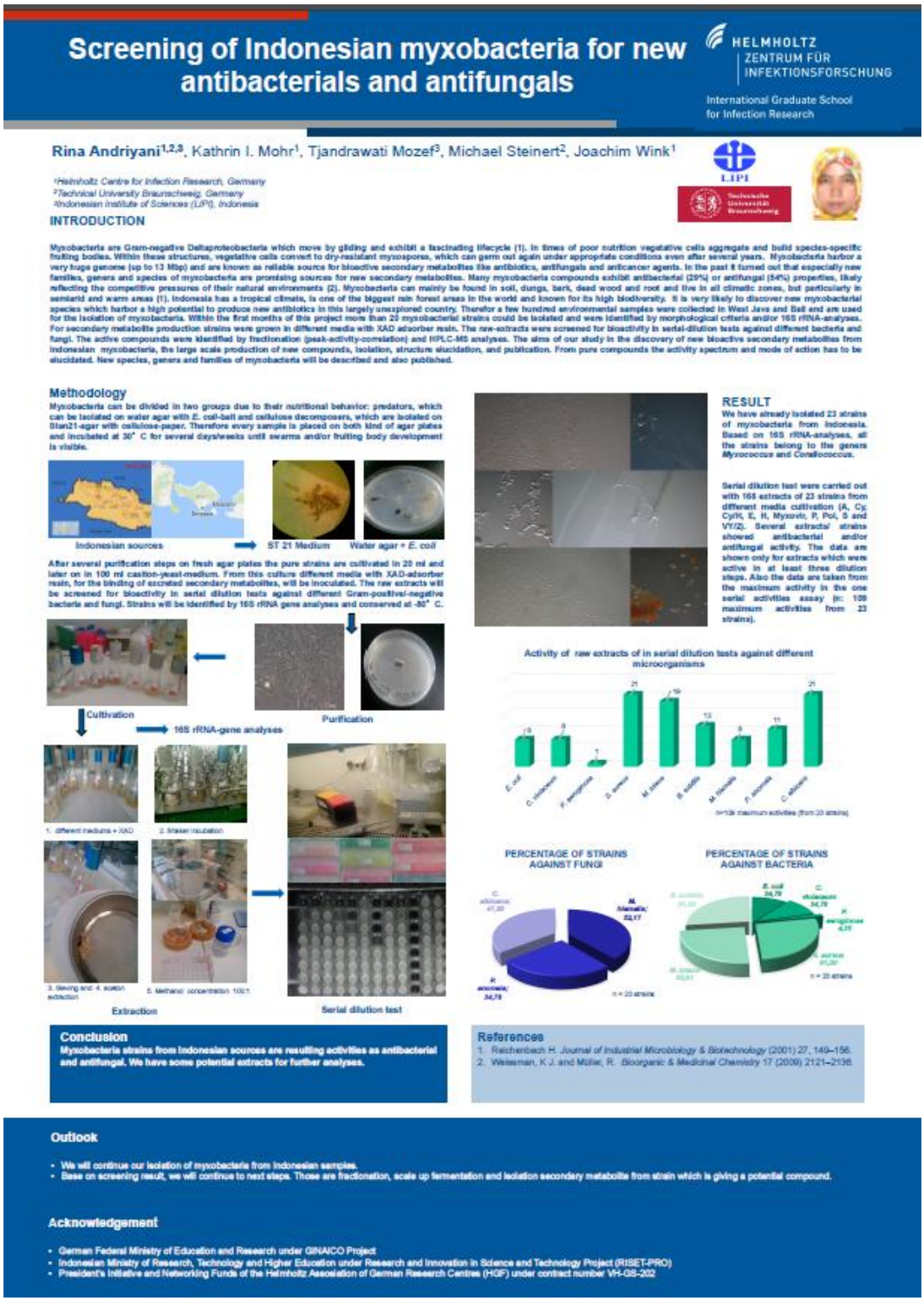
9. Sequence 1: MxGP1963KM***
   Sequence 2: Mxx1929-1KR***

Score	Expect	Identities	Gaps	Strand
1672 bits(905)	0.0	909/911(99%)	0/911(0%)	Plus/Plus

```
Query  1    GTCTCCGTCGCCGGACCGTCGGCGATGACGTCACCCTTCTTCACCCGGTCGCCCTTGCTG  60
Sbjct  22   ............................................................  81

Query  61   ATGATGGGCTTCTGGTTGAGGCAGGTGTTCTGGTTGGAGCGCTGGTACTTGAGCAGGTTG  120
Sbjct  82   ............................................................  141

Query  121  TAGATGTCGACCTCGCTCGTGACGTCGCTCAGCGAGGCCGGCACGTCCGCCTTCACCACG  180
Sbjct  142  ............................................................  201

Query  181  ATGCGGCCGGCGTCCACGCTCTCCACGATGCCGTCGCGCCGGGCCACGCAGGTGACGCCG  240
Sbjct  202  ............................................................  261

Query  241  GAGTCGCGGGCGACGATCGACTCGATGCCCGTGCCCACCAGCGGCGCCGCGGTTGGCGTCGTCGTTC  300
Sbjct  262  ............................................................  321

Query  301  AGCGGAACGGCCTGGCGCTGCATGTTGGAGCCCATGAGCGCGCGGTTGGCGTCGTCGTTC  360
Sbjct  322  ............................................................  381

Query  361  TCCAGGAACGGGATGAGCGACGCGGCCACCGACACCAGCTGCTTCGGGGACACGTCCATC  420
Sbjct  382  ............................................................  441

Query  421  AGGTCCACGTCCTCGGCGCGCGCCTGGACGAACTCACCCCCGCGGCGGCTGGACACCAGC  480
Sbjct  442  ............................................................  501

Query  481  GCGTTGGCGAACTTGCCCTTCTTGTCCGTCTCCGCGTTCGCCTGGGCGATGGTGTGCTTC  540
Sbjct  502  ............................................................  561

Query  541  TCCTCCTCGAGCGCGGAGTAGAAGGCCACGTCGTTGGTCACCACGCCGGCGTCGACCTTG  600
Sbjct  562  ...........................................A................  621

Query  601  CGGTAGGGCGTCTCCACGAAGCCGAACTCGTTGACGCGCGCGTAGGTGGACAGCGACGCG  660
Sbjct  622  ............................................................  681

Query  661  ATGAGGCCGATGTTCGGACCTTCCGGCGTCTCGATGGGGCAGATGCGGCCGTAGTGCGTC  720
Sbjct  682  ............................................................  741

Query  721  GGGTGCACGTCGCGCACCTCGAAGCCGGCGCGCTCGCGGGTGAGGCCGCCGGGCCCCAGC  780
Sbjct  742  ............................................................  801

Query  781  GCGGACAGGCGCCGCTTGTGGGTGACTTCCGACAGGGGGTTCGTCTGGTCCATGAACTGC  840
Sbjct  802  ............................................................  861

Query  841  GACAGCTGGCTGGACCCGAAGAACTCCTTGATGACCGCCGTGACGGGCTTGGCGTTGATG  900
Sbjct  862  ...........................................C................  921

Query  901  AGGTCGTGCGG  911
Sbjct  922  ...........  932
```

Supplement 7. List of posters

1.



Screening of Indonesian myxobacteria for new antibacterials and antifungals

HELMHOLTZ ZENTRUM FÜR INFEKTIONSFORSCHUNG

International Graduate School for Infection Research

Rina Andriyani[1,2,3], Kathrin I. Mohr[1], Tjandrawati Mozef[3], Michael Steinert[2], Joachim Wink[1]

[1]Helmholtz Centre for Infection Research, Germany
[2]Technical University Braunschweig, Germany
[3]Indonesian Institute of Sciences (LIPI), Indonesia

INTRODUCTION

Myxobacteria are Gram-negative Deltaproteobacteria which move by gliding and exhibit a fascinating lifecycle (1). In times of poor nutrition vegetative cells aggregate and build species-specific fruiting bodies. Within these structures, vegetative cells convert to dry-resistant myxospores, which can germ out again under appropriate conditions even after several years. Myxobacteria harbor a very huge genome (up to 13 Mbp) and are known as reliable source for bioactive secondary metabolites like antibiotics, antifungals and anticancer agents. In the past it turned out that especially new families, genera and species of myxobacteria are promising sources for new secondary metabolites. Many myxobacteria compounds exhibit antibacterial (29%) or antifungal (54%) properties, likely reflecting the competitive pressures of their natural environments (2). Myxobacteria can mainly be found in soil, dungs, bark, dead wood and root and live in all climatic zones, but particularly in semiarid and warm areas (1). Indonesia has a tropical climate, is one of the biggest rain forest areas in the world and known for its high biodiversity. It is very likely to discover new myxobacterial species which harbor a high potential to produce new antibiotics in this largely unexplored country. Therefor a few hundred environmental samples were collected in West Java and Bali and are used for the isolation of myxobacteria. Within the first months of this project more than 20 myxobacterial strains could be isolated and were identified by morphological criteria and/or 16S rRNA-analyses. For secondary metabolite production strains were grown in different media with XAD adsorber resin. The raw-extracts were screened for bioactivity in serial-dilution tests against different bacteria and fungi. The active compounds were identified by fractionation (peak-activity-correlation) and HPLC-MS analyses. The aims of our study in the discovery of new bioactive secondary metabolites from Indonesian myxobacteria, the large scale production of new compounds, isolation, structure elucidation, and publication. From pure compounds the activity spectrum and mode of action has to be elucidated. New species, genera and families of myxobacteria will be described and also published.

Methodology

Myxobacteria can be divided in two groups due to their nutritional behavior: predators, which can be isolated on water agar with *E. coli*-bait and cellulose decomposers, which are isolated on Stan21-agar with cellulose-paper. Therefore every sample is placed on both kind of agar plates and incubated at 30° C for several days/weeks until swarms and/or fruiting body development is visible.

Indonesian sources → ST 21 Medium, Water agar + *E. coli*

After several purification steps on fresh agar plates the pure strains are cultivated in 20 ml and later on in 100 ml casiton-yeast-medium. From this culture different media with XAD-adsorber resin, for the binding of excreted secondary metabolites, will be inoculated. The raw extracts will be screened for bioactivity in serial dilution tests against different Gram-positive/-negative bacteria and fungi. Strains will be identified by 16S rRNA gene analyses and conserved at -80° C.

Cultivation → Purification

16S rRNA-gene analyses

1. different mediums + XAD 2. Shaker incubation

3. Mixing and 4. solvent extraction 5. Methanol concentration 100:1

Extraction → Serial dilution test

RESULT

We have already isolated 23 strains of myxobacteria from Indonesia. Based on 16S rRNA-analyses, all the strains belong to the genera *Myxococcus* and *Corallococcus*.

Serial dilution test were carried out with 168 extracts of 23 strains from different media cultivation (A, Cy, CyH, E, H, Myxovit, P, Pol, S and VY/2). Several extracts/ strains showed antibacterial and/or antifungal activity. The data are shown only for extracts which were active in at least three dilution steps. Also the data are taken from the maximum activity in the one serial activities assay (n: 109 maximum activities from 23 strains).

Activity of raw extracts of in serial dilution tests against different microorganisms

n=109 maximum activities (from 23 strains)

PERCENTAGE OF STRAINS AGAINST FUNGI

n = 23 strains

PERCENTAGE OF STRAINS AGAINST BACTERIA

Conclusion

Myxobacteria strains from Indonesian sources are resulting activities as antibacterial and antifungal. We have some potential extracts for further analyses.

References

1. Reichenbach H. *Journal of Industrial Microbiology & Biotechnology* (2001) 27, 149–156.
2. Weissman, K.J. and Müller, R. *Bioorganic & Medicinal Chemistry* 17 (2009) 2121–2136

Outlook

- We will continue our isolation of myxobacteria from Indonesian samples.
- Base on screening result, we will continue to next steps. Those are fractionation, scale up fermentation and isolation secondary metabolite from strain which is giving a potential compound.

Acknowledgement

- German Federal Ministry of Education and Research under GINAICO Project
- Indonesian Ministry of Research, Technology and Higher Education under Research and Innovation in Science and Technology Project (RISET-PRO)
- President's Initiative and Networking Funds of the Helmholtz Association of German Research Centres (HGF) under contract number VH-GS-202

2.

Compounds from *Myxococcus* sp. (MxGP53TM) a strain isolated from Indonesian soil

HELMHOLTZ ZENTRUM FÜR INFEKTIONSFORSCHUNG

International Graduate School for Infection Research

Rina Andriyani[1,2,3], Enge Sudarman[1], Kathrin I. Mohr[1], Tjandrawati Mozef[3], Michael Steinert[2], Joachim Wink[1]

[1]Helmholtz Centre for Infection Research, Germany, [2]Technical University Braunschweig, Germany, [3]Indonesian Institute of Sciences (LIPI) – Indonesia.

Introduction

Myxobacteria are Gram-negative and move by gliding. They are unique in the bacterial kingdom due to their ability to produce species-specific fruiting bodies with dry resistant myxospores (1). The most interesting feature is their ability to produce numerous secondary metabolites, many of them with bioactivity (2, 3). With one exception, myxobacteria are strictly aerobic. They can be isolated from soil, dung, compost, bark, dead trees and roots, but also genera from marine environments have been described. Myxobacteria can be found in all climatic zones, particularly in semiarid and warm areas. In the past it turned out that especially new species, genera and families of myxobacteria are promising sources to detect new secondary metabolites.

Indonesia with its manifold biodiversity has a huge potential for the isolation of new uncommon myxobacteria, which are potential sources for novel natural products.

The aim of our study is the isolation and structure elucidation of unknown compounds from myxobacteria isolated from Indonesian samples. From strain MxGP53TM two new compounds showing activity against Gram-positive bacteria could be identified.

Result

The orange colored strain MxGP53TM (Figure 2) was isolated from a soil sample, collected in Gunung Padang (West Java, Indonesia, Figure 1). After DNA-extraction, PCR with primer specific for the 16S rRNA-gene (f27, r1525), and sequencing, comparison with sequences of the public database NCBI revealed 99.4% similarity to the type strain of *Myxococcus stipitatus* (family Myxococcaceae, suborder Cystobacterineae).

The strain was cultivated in 10 different media with XAD-adsorber resin for one week at 30 ° C. Raw extracts were tested against different Gram+ and – bacteria, yeasts and a filamentous fungi in serial dilution tests. Thereby, medium VY/2, containing yeast *Saccharomyces cerevisiae*, showed the best activity against a couple of test organisms (Table 1; Figure 4).

Figure 2. MxGP53TM: Morphology of cells, fruiting bodies on agar and grown in medium CY/H

Methodology

Figure 1. Schedule: From environmental sample to bioactive compounds

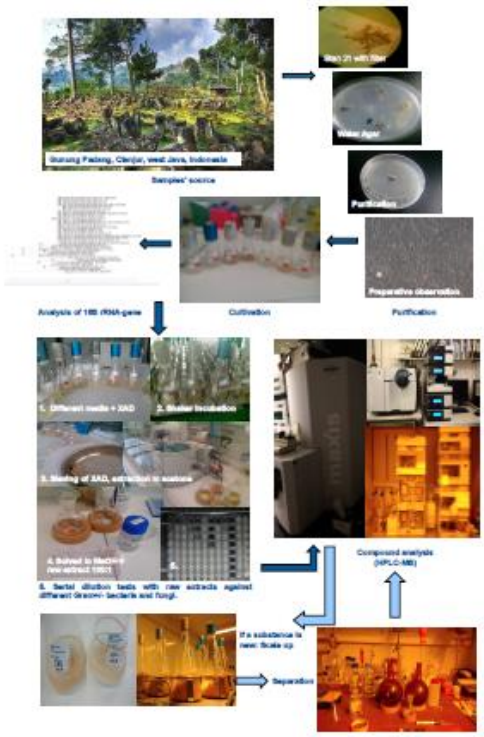


Table 1: Results of serial dilution tests of raw extract MxGP53TM grown in VY/2 medium against different test organisms: the higher the letter the stronger the inhibition of the tested germ (A < B < C...)

E. coli DSM 1116	*C. violaceum* DSM 30191	*P. aeruginosa* PA14	*S. aureus* Newman	*M. luteus* DSM 1790	*B. subtilis* DSM 10
F	H	E	H	H	H

For peak-activity correlation, the raw extract of VY/2-medium was fractionated by analytical HPLC (DAD-UV detector combined with high resolution ESI mass spectrometry; HRESIMS) and inoculated with the former inhibited test organisms (Figure 5). After incubation, inhibited wells could be correlated to specific UV-spectra and retention time. Besides known myxalamids, myxochelin B and some new derivatives of DK-xantheres, two new compounds could be identified, whereas one of the compounds has been proven to inhibit *S. aureus* at retention time 14.49 min. The mass is known, but confidential.

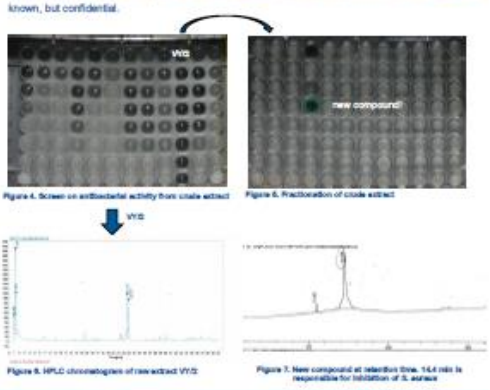


Figure 4. Screen on antibacterial activity from crude extract

Figure 6. Fractionation of crude extract

Figure 5. HPLC chromatogram of raw extract VY/2

Figure 7. New compound at retention time, 14.4 min is responsible for inhibition of *S. aureus*

Conclusion

- Sequencing of 16S rRNA–gene revealed 99.4% similarity of MxGP53TM to *Myxococcus stipitatus*.
- Cultivation of MxGP53TM in 10 different media revealed VY/2 as the best for production of secondary metabolites
- Raw extract of VY/2-medium was active against Gram-positive as well as against Gram-negative bacteria
- Two novel compounds with identified masses (confidential) as well as myxalamids and some new derivatives of DK-xantheres and myxochelin B, are produced by strain MxGP53TM.

Outlook

Optimization of production of the two new compounds by testing different C- and N-sources, temperatures, pHs, cultivation times and aeration.

If productivity is satisfactory: large-scale cultivation to produce sufficient amounts for structure elucidation (NMR) and minimal inhibition concentration-tests (MIC) against clinical relevant germs.

References

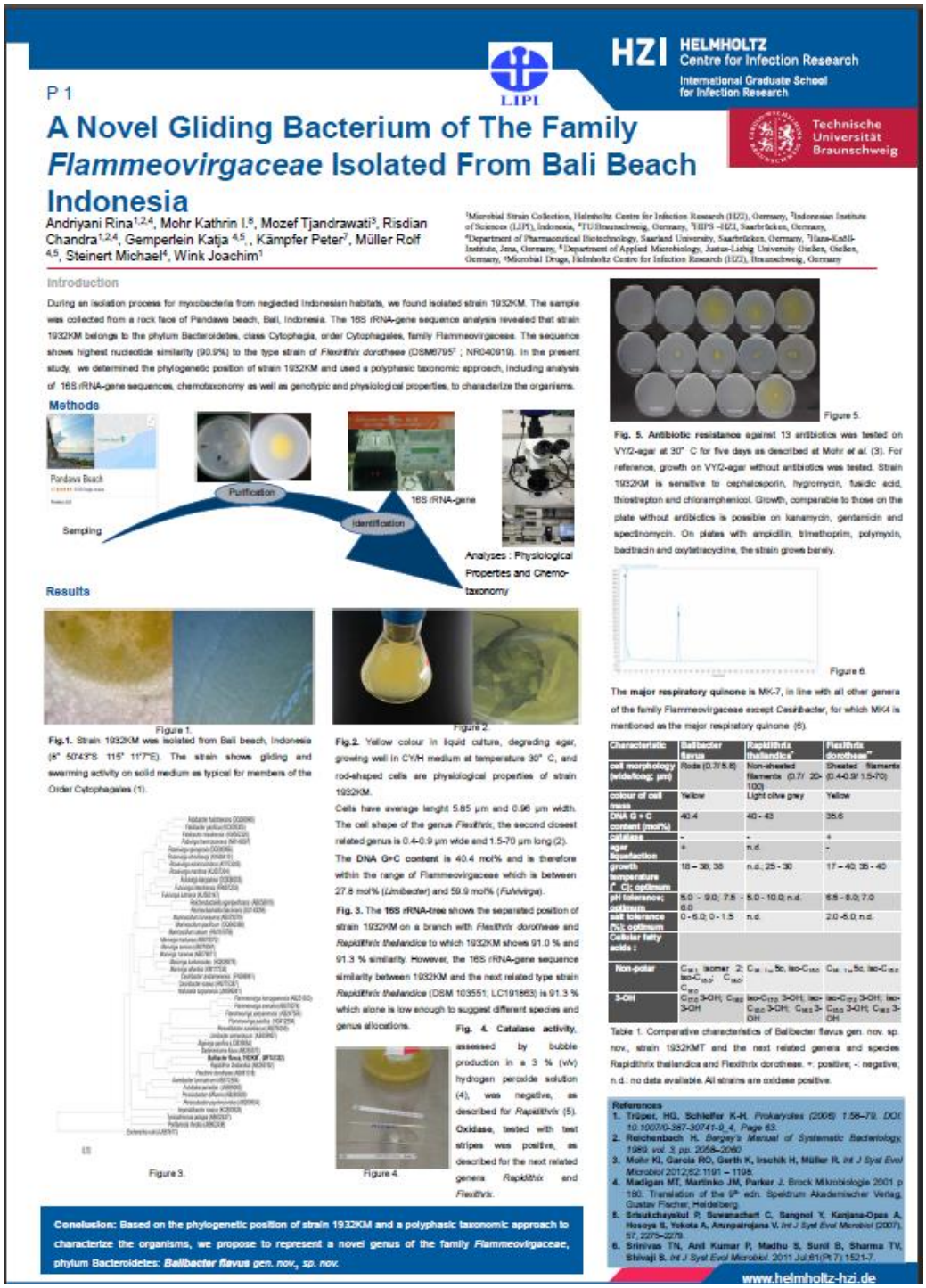
1. Reichenbach H. (2001), J Industrial Microbiol & Biotech, 27, 149–156.
2. Weissman KJ, Müller R. (2010) Nat Prod Rep 27:1276-1295.
3. Herrmann J, Abou Fayad A, Müller R. (2017) Nat Prod Rep 34: 135-160

Acknowledgement

- German Federal Ministry of Education and Research under GINAICO Project
- Indonesian Ministry of Research, Technology and Higher Education under Research and Innovation in Science and Technology Project (RISET-PRO)
- President's Initiative and Networking Funds of the Helmholtz Association of German Research Centres (HGF) under contract number VH-GS-202

Poster Presentation for The 10th PhD Symposium, International Graduate School for Infection Research HZI on December 7th, 2017 in Inhoffenstraße 7, 38124 Braunschweig, Germany

3.



P 1

A Novel Gliding Bacterium of The Family *Flammeovirgaceae* Isolated From Bali Beach Indonesia

Andriyani Rina[1,2,4], Mohr Kathrin I.[6], Mozef Tjandrawati[5], Risdian Chandra[1,2,4], Gemperlein Katja [4,5], , Kämpfer Peter[7], Müller Rolf [4,5], Steinert Michael[4], Wink Joachim[1]

[1]Microbial Strain Collection, Helmholtz Centre for Infection Research (HZI), Germany, [2]Indonesian Institute of Sciences (LIPI), Indonesia, [3]TU Braunschweig, Germany, [4]HIPS –HZI, Saarbrücken, Germany, [5]Department of Pharmaceutical Biotechnology, Saarland University, Saarbrücken, Germany, [6]Hans-Knöll-Institute, Jena, Germany, [7]Department of Applied Microbiology, Justus-Liebig University Gießen, Gießen, Germany, [8]Microbial Drugs, Helmholtz Centre for Infection Research (HZI), Braunschweig, Germany

Introduction

During an isolation process for myxobacteria from neglected Indonesian habitats, we found isolated strain 1932KM. The sample was collected from a rock face of Pandawa beach, Bali, Indonesia. The 16S rRNA-gene sequence analysis revealed that strain 1932KM belongs to the phylum Bacteroidetes, class Cytophagia, order Cytophagales, family Flammeovirgaceae. The sequence shows highest nucleotide similarity (90.9%) to the type strain of *Flexithrix dorotheae* (DSM6797 ; NR040919). In the present study, we determined the phylogenetic position of strain 1932KM and used a polyphasic taxonomic approach, including analysis of 16S rRNA-gene sequences, chemotaxonomy as well as genotypic and physiological properties, to characterize the organisms.

Methods

Results

Figure 1.

Fig.1. Strain 1932KM was isolated from Bali beach, Indonesia (8° 50'43"S 115° 11'7"E). The strain shows gliding and swarming activity on solid medium as typical for members of the Order Cytophagales (1).

Figure 2.

Fig.2. Yellow colour in liquid culture, degrading agar, growing well in CY/H medium at temperature 30° C, and rod-shaped cells are physiological properties of strain 1932KM.

Cells have average lenght 5.85 µm and 0.98 µm width. The cell shape of the genus *Flexithrix*, the second closest related genus is 0.4-0.9 µm wide and 1.5-70 µm long (2).

The **DNA G+C content** is 40.4 mol% and is therefore within the range of Flammeovirgaceae which is between 27.8 mol% (*Limibacter*) and 59.9 mol% (*Fulvivirga*).

Figure 3.

Fig. 3. The **16S rRNA-tree** shows the separated position of strain 1932KM on a branch with *Flexithrix dorotheae* and *Rapidithrix thailandica* to which 1932KM shows 91.0 % and 91.3 % similarity. However, the 16S rRNA-gene sequence similarity between 1932KM and the next related type strain *Rapidithrix thailandica* (DSM 103551; LC191883) is 91.3 % which alone is low enough to suggest different species and genus allocations.

Figure 4.

Fig. 4. Catalase activity, assessed by bubble production in a 3 % (w/v) hydrogen peroxide solution (4), was negative, as described for *Rapidithrix* (5). **Oxidase**, tested with test stripes was positive, as described for the next related genera *Rapidithrix* and *Flexithrix*.

Figure 5.

Fig. 5. Antibiotic resistance against 13 antibiotics was tested on VY/2-agar at 30° C for five days as described at Mohr *et al.* (3). For reference, growth on VY/2-agar without antibiotics was tested. Strain 1932KM is sensitive to cephalosporin, hygromycin, fusidic acid, thiostrepton and chloramphenicol. Growth, comparable to those on the plate without antibiotics is possible on kanamycin, gentamicin and spectinomycin. On plates with ampicillin, trimethoprim, polymyxin, bacitracin and oxytetracycline, the strain grows barely.

Figure 6.

The major respiratory quinone is MK-7, in line with all other genera of the family Flammeovirgaceae except *Cesiribacter*, for which MK4 is mentioned as the major respiratory quinone (6).

Characteristic	*Balibacter flavus*	*Rapidithrix thailandica*[1]	*Flexithrix dorotheae*[2]
cell morphology (wide/long; µm)	Rods (0.7/ 5.6)	Non-sheated filaments (0.7/ 20-100)	Sheated filaments (0.44-0.9/ 1.5-70)
colour of cell mass	Yellow	Light olive grey	Yellow
DNA G + C content (mol%)	40.4	40 - 43	35.6
catalase	-	-	+
agar hydrolysis	+	n.d.	-
growth temperature (° C); optimum	18 – 38; 38	n.d.; 25 - 30	17 – 40; 35 - 40
pH tolerance; optimum	5.0 - 9.0; 7.5 - 8.0	5.0 - 10.0; n.d.	6.5 - 8.0; 7.0
salt tolerance (%); optimum	0 - 6.0; 0 - 1.5	n.d.	2.0 -5.0; n.d.
Cellular fatty acids :			
Non-polar	$C_{16:1}$ isomer 2; $C_{16:1\omega5c}$, iso-$C_{15:0}$, iso-$C_{16:0}$	$C_{16:1\omega5c}$, iso-$C_{15:0}$	$C_{16:1\omega5c}$, iso-$C_{15:0}$
3-OH	$C_{12:0}$ 3-OH; $C_{16:0}$ 3-OH	iso-$C_{17:0}$ 3-OH; iso-$C_{15:0}$ 3-OH; $C_{16:0}$ 3-OH	iso-$C_{17:0}$ 3-OH; iso-$C_{15:0}$ 3-OH; $C_{16:0}$ 3-OH

Table 1. Comparative characteristics of Balibacter flavus gen. nov. sp. nov., strain 1932KMT and the next related genera and species Rapidithrix thailandica and Flexithrix dorotheae. +: positive; -: negative; n.d.: no data available. All strains are oxidase positive.

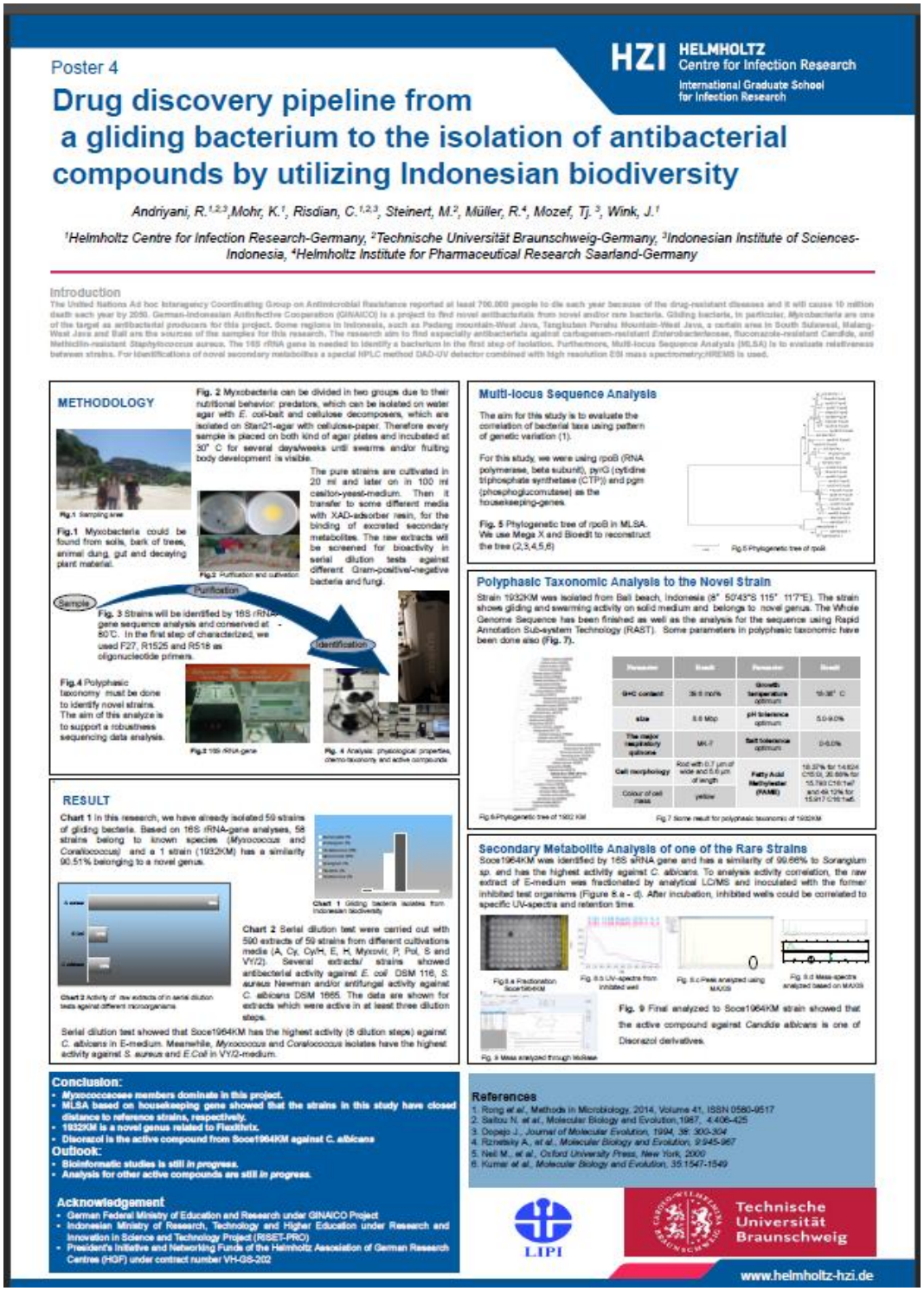
Conclusion: Based on the phylogenetic position of strain 1932KM and a polyphasic taxonomic approach to characterize the organisms, we propose to represent a novel genus of the family *Flammeovirgaceae*, phylum Bacteroidetes: *Balibacter flavus gen. nov., sp. nov.*

References

1. Trüper, HG, Schleifer K-H. *Prokaryotes* (2006) *1:58–72. DOI: 10.1007/0-387-30741-9_4, Page 63.*
2. Reichenbach H. *Bergey's Manual of Systematic Bacteriology, 1989, vol. 3, pp. 2058–2060*
3. Mohr KI, Garcia RO, Gerth K, Irschik H, Müller R. *Int J Syst Evol Microbiol* 2012;62:1191 – 1196.
4. Madigan MT, Martinko JM, Parker J. Brock Mikrobiologie 2001 p 180. Translation of the 9th edn. Spektrum Akademischer Verlag, Gustav Fischer, Heidelberg.
5. Siriukchayakul P, Sowanachan C, Sangnol Y, Kanjana-Opas A, Hosoya S, Yokota A, Amnatpraipana V. *Int J Syst Evol Microbiol* (2007), 57, 2278–2279.
6. Srinivas TN, Anil Kumar P, Madhu S, Sunil B, Sharma TV, Shivaji S. *Int J Syst Evol Microbiol.* 2011 Jul;61(Pt 7):1521-7.

www.helmholtz-hzi.de

4.



Poster 4

Drug discovery pipeline from a gliding bacterium to the isolation of antibacterial compounds by utilizing Indonesian biodiversity

HZI HELMHOLTZ Centre for Infection Research
International Graduate School for Infection Research

Andriyani, R.[1,2,3], Mohr, K.[1], Risdian, C.[1,2,3], Steinert, M.[2], Müller, R.[4], Mozef, Tj.[3], Wink, J.[1]

[1]Helmholtz Centre for Infection Research-Germany, [2]Technische Universität Braunschweig-Germany, [3]Indonesian Institute of Sciences-Indonesia, [4]Helmholtz Institute for Pharmaceutical Research Saarland-Germany

Introduction

The United Nations Ad hoc Interagency Coordinating Group on Antimicrobial Resistance reported at least 700.000 people to die each year because of the drug-resistant diseases and it will cause 10 million death each year by 2050. German-Indonesian Antiinfective Cooperation (GINAICO) is a project to find novel antibacterials from novel and/or rare bacteria. Gliding bacteria, in particular, Myxobacteria are one of the target as antibacterial producers for this project. Some regions in Indonesia, such as Padang mountain-West Java, Tangkuban Perahu Mountain-West Java, a certain area in South Sulawesi, Malang-West Java and Bali are the sources of the samples for this research. The research aim to find especially antibacterials against carbapenem-resistant Enterobacteriaceae, fluconazole-resistant Candida, and Methicilin-resistant Staphylococcus aureus. The 16S rRNA gene is needed to identify a bacterium in the first step of isolation. Furthermore, Multi-locus Sequence Analysis (MLSA) is to evaluate relatiness between strains. For identifications of novel secondary metabolites a special HPLC method DAD-UV detector combined with high resolution ESI mass spectrometry(HRESMS is used.

METHODOLOGY

Fig.1 Sampling area

Fig.1 Myxobacteria could be found from soils, bark of trees, animal dung, gut and decaying plant material.

Fig. 2 Myxobacteria can be divided in two groups due to their nutritional behavior: predators, which can be isolated on water agar with *E. coli*-bait and cellulose decomposers, which are isolated on Stan21-agar with cellulose-paper. Therefore every sample is placed on both kind of agar plates and incubated at 30° C for several days/weeks until swarms and/or fruiting body development is visible.

The pure strains are cultivated in 20 ml and later on in 100 ml casiton-yeast-medium. Then it transfer to some different media with XAD-adsorber resin, for the binding of excreted secondary metabolites. The raw extracts will be screened for bioactivity in serial dilution tests against different Gram-positive/-negative bacteria and fungi.

Fig.2 Purification and cultivation

Sample → Purification → Identification

Fig. 3 Strains will be identified by 16S rRNA gene sequence analysis and conserved at -80°C. In the first step of characterized, we used F27, R1525 and R518 as oligonucleotide primers.

Fig.4 Polyphasic taxonomy must be done to identify novel strains. The aim of this analyze is to support a robustness sequencing data analysis.

Fig.3 16S rRNA gene

Fig. 4 Analysis: physiological properties, chemo-taxonomy and active compounds

Multi-locus Sequence Analysis

The aim for this study is to evaluate the correlation of bacterial taxa using pattern of genetic variation (1).

For this study, we were using rpoB (RNA polymerase, beta subunit), pyrG (cytidine triphosphate synthetase (CTP)) and pgm (phosphoglucomutase) as the housekeeping-genes.

Fig. 5 Phylogenetic tree of rpoB in MLSA. We use Mega X and Bioedit to reconstruct the tree (2,3,4,5,6)

Fig.5 Phylogenetic tree of rpoB

Polyphasic Taxonomic Analysis to the Novel Strain

Strain 1932KM was isolated from Bali beach, Indonesia (8° 50'43"S 115° 117"E). The strain shows gliding and swarming activity on solid medium and belongs to novel genus. The Whole Genome Sequence has been finished as well as the analysis for the sequence using Rapid Annotation Sub-system Technology (RAST). Some parameters in polyphasic taxonomic have been done also (Fig. 7).

Parameter	Result	Parameter	Result
GHC content	39.6 mol%	Growth temperature optimum	18-36° C
size	8.6 Mbp	pH tolerance optimum	5.0-9.0%
The major respiratory quinone	MK-7	Salt tolerance optimum	0-5.0%
Cell morphology	Rod with 0.7 µm of wide and 5.6 µm of length	Fatty Acid Methylester (FAME)	18.37% for 14:0/24 C15:0, 20.69% for 15.783 C16:1w7 and 49.10% for 15.917 C16:1w6.
Colour of cell mass	yellow		

Fig.6 Phylogenetic tree of 1932 KM

Fig.7 Some result for polyphasic taxonomic of 1932KM

RESULT

Chart 1 In this research, we have already isolated 59 strains of gliding bacteria. Based on 16S rRNA-gene analyses, 58 strains belong to known species (*Myxococcus* and *Corallococcus*) and a 1 strain (1932KM) has a similarity 90.51% belonging to a novel genus.

Chart 1 Gliding bacteria isolates from Indonesian biodiversity

Chart 2 Activity of raw extracts of in serial dilution tests against different microorganisms

Chart 2 Serial dilution test were carried out with 590 extracts of 59 strains from different cultivations media (A, Cy, CyH, E, H, Myxovir, P, Pol, S and VY/2). Several extracts/ strains showed antibacterial activity against *E. coli* DSM 116, *S. aureus* Newman and/or antifungal activity against *C. albicans* DSM 1665. The data are shown for extracts which were active in at least three dilution steps.

Serial dilution test showed that Soce1964KM has the highest activity (8 dilution steps) against *C. albicans* in E-medium. Meanwhile, *Myxococcus* and *Coralococcus* isolates have the highest activity against *S. aureus* and *E.Coli* in VY/2-medium.

Secondary Metabolite Analysis of one of the Rare Strains

Socs1964KM was identified by 16S rRNA gene and has a similarity of 99.66% to *Sorangium* sp. and has the highest activity against *C. albicans*. To analysis activity correlation, the raw extract of E-medium was fractionated by analytical LC/MS and inoculated with the former inhibited test organisms (Figure 8.a - d). After incubation, inhibited wells could be correlated to specific UV-spectra and retention time.

Fig.8.a Fractionation Socs1964KM

Fig. 8.b UV-spectra from inhibited well

Fig. 8.c Peak analyzed using MAXIS

Fig. 8.d Mass-spectra analyzed based on MAXIS

Fig. 9 Mass analyzed through MoNaa

Fig. 9 Final analyzed to Socs1964KM strain showed that the active compound against Candida albicans is one of Discrazol derivatives.

Conclusion:

- Myxococcaceae members dominate in this project.
- MLSA based on housekeeping gene showed that the strains in this study have closed distance to reference strains, respectively.
- 1932KM is a novel genus related to Flexithrix.
- Discrazol is the active compound from Socs1964KM against C. albicans

Outlook:

- Bioinformatic studies is still *in progress*.
- Analysis for other active compounds are still *in progress*.

Acknowledgement

- German Federal Ministry of Education and Research under GINAICO Project
- Indonesian Ministry of Research, Technology and Higher Education under Research and Innovation in Science and Technology Project (RISET-PRO)
- President's Initiative and Networking Funds of the Helmholtz Association of German Research Centres (HGF) under contract number VH-GS-202

References

1. Rong *et al.*, Methods in Microbiology, 2014, Volume 41, ISSN 0580-9517
2. Saitou N. *et al.*, Molecular Biology and Evolution,1987, 4:406-425
3. Dopejo J., *Journal of Molecular Evolution, 1994, 38: 300-304*
4. Rzhetsky A., *et al., Molecular Biology and Evolution, 9:945-967*
5. Nei M., *et al., Oxford University Press, New York, 2000*
6. Kumar *et al., Molecular Biology and Evolution, 35:1547-1549*

LIPI

Technische Universität Braunschweig

www.helmholtz-hzi.de